University of Alberta

A New Curvature-Compensation Technique for Bandgap Voltage Reference

by

Yiu Fai Chan

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

Department of Electrical and Computer Engineering

Edmonton, Alberta
Spring 1998

National Library of Canada

Acquisitions and Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada

Acquisitions et services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-28924-9

Canada

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **A New Curvature-Compensation Technique for Bandgap Voltage Reference** submitted by Yiu Fai Chan in partial fulfillment of the requirements for the degree of **Master of Science**.

I. Filanovsky

Dr. Igor M. Filanovsky

John C. Salmon

Dr. John Salmon

Doug. Gingrich

Dr. Douglas M. Gingrich

Date: Dec 1, 1997.

To my wife, Denise, my sisters, Mandy, Abby, Valerie, my parents and my parents-in-law

Abstract

This thesis introduces a new curvature-compensation technique for bandgap voltage reference. This technique utilizes a new idea to linearize the temperature dependence of base-emitter voltage of the bipolar transistor in bandgap reference design. The linearization is obtained by supplying the collector current which dependence on absolute temperature is described by a polynomial $AT^3 + BT^4$. The temperature dependence of the current is realized via a special translinear circuit fed by two currents from an auxiliary bandgap reference. One of this current is independent on temperature, and another is proportional to the absolute temperature (PTAT). The theoretical derivation shows that the new technique can reduce the overall error of a bandgap reference to be within 60 μV for the temperature range of -55 ^oC to 180 ^oC.

The basic circuit topology of the new bandgap reference design is presented. Two bandgap reference designs are derived from this topology. One of them is designed for use with a 5 V supply voltage while the other for a 3 V supply voltage. The simulation results show that the 5 V design can produce a reference with 5 ppm/^oC and the 3 V design can produce a 9 ppm/^oC reference in the same temperature range of -55 ^oC to 180 ^oC. The layouts for the two designs were done and the physical effects in the layouts were back-annotated. The simulation results of the two layouts also demonstrated a similar performance as those in the transistor level. All of these results suggest that the new technique can produce a highly accurate bandgap reference design which can be used in many electronic systems.

Acknowledgements

I would like to thank my supervisor, Dr. Filanovsky, for his guidance and support throughout my graduate work. I also would like to thank the Government of Alberta and Macau Foundation for financial support.

Contents

1 Introduction 1

1.1 Voltage Reference Implementation . . . 3

1.2 Monolithic Circuit Technologies . . . 5

1.3 Thesis Content . . . 6

2 An Overview of Temperature-compensated Voltage References 9

2.1 Introduction . . . 9

2.2 Zener Voltage Reference . . . 10

2.3 Bandgap Voltage Reference . . . 12

2.4 First-order Compensated Designs . . . 16

2.4.1 Widlar Bandgap Reference . . . 17

2.4.2 Kuijk Bandgap Reference . . . 19

2.4.3 Brokaw Bandgap Reference . . . 21

2.4.4 Discussions . . . 23

2.5 Other Technologies . . . 24

2.6 Curvature-Compensated Bandgap Reference . . . 28

2.7 Conclusion . . . 31

3 Translinear Circuits 33

3.1 Introduction . . . 33

3.2 Bipolar Translinear Principle . . . 35

3.3 Practical Circuits . . . 38

3.3.1 One-quadrant Squarer . . . 39

3.3.2 Two-quadrant Squarer . . . 42

3.3.3 Rooting and Geometric-mean Circuit . . . 45

3.3.4 Vector Summation Circuit 45
3.4 Non-ideal Behaviors 47
3.5 MOS Translinear Principle 49
3.6 Conclusion 52

4 A New Compensation Technique **53**
4.1 Introduction 53
4.2 V_{BE} Linearization 54
4.3 The Bandgap Reference Circuit Configuration 57
4.3.1 5 V Supply Bandgap Reference Circuit 61
4.3.2 3 V Supply Bandgap Reference Circuit 68
4.4 Conclusion 73

5 Simulation Tools and Results **76**
5.1 Introduction 76
5.2 A Brief History in Circuit Simulation 76
5.3 The Features in Spectre 78
5.4 Simulation Results of the 5 V Design 80
5.5 Simulation Results of the 3 V Design 82
5.6 Conclusion 83

6 Layout Design **92**
6.1 Introduction 92
6.2 Layout Considerations 94
6.3 Simulation Results 98
6.4 Conclusion 99

7 Conclusion **106**

Bibliography **109**

List of Tables

4.1 The Design Parameters of the 5 V Supply Bandgap reference 68

4.2 The Design Parameters of the 3 V Supply Bandgap reference 73

5.1 The Performance Parameters of the 5 V Supply Bandgap reference . 82

5.2 The Performance Parameters of the 3 V Supply Bandgap reference . 84

List of Figures

2.1 A Zener Voltage Reference Circuit 11

2.2 A Symbolic Representation of Bandgap Voltage Reference 13

2.3 The Nonlinearity $(\eta - m)(k/q)\{T - T_r - T\ln(T/T_r)\}$ of $V_{BE}(T)$ vs. T/T_r 16

2.4 Widlar Bandgap Voltage Reference 18

2.5 Kuijk Bandgap Voltage Reference 20

2.6 Brokaw Bandgap Voltage Reference 22

2.7 A Typical CMOS Bandgap Reference (with n-well process) 26

2.8 A CMOS Voltage Reference with MOS Transistors Operating in Weak Inversion 27

3.1 A Generalized Bipolar Translinear Loop 36

3.2 One-quadrant TL Squarer/Divider [26] 39

3.3 Power-n Circuit 40

3.4 Circuit Generating Currents Proportional to T^3 and T^4 for 5 V Supply Voltage 41

3.5 Circuit Generating Currents Proportional to T^3 and T^4 for 3 V Supply Voltage 42

3.6 Absolute-value Circuits [26] 43

3.7 Two-quadrant TL Squarer [26] 44

3.8 A TL Square-rooting and Geometric-mean Circuit [26] 45

3.9 A TL Vector Summation Circuit [26] 46

3.10 A Generalized MOS Translinear Loop 51

4.1 The Residual Error Resulted from the New Linearization Technique . 57

4.2 The Achievable Residual Error of the New Design Technique 58

4.3 Bandgap Reference Circuit Configuration 59
4.4 Bandgap Reference Circuit for 5 V Supply Voltage 74
4.5 Bandgap Reference Circuit for 3 V Supply Voltage 75

5.1 The Reference Voltage of the 5 V Design versus Temperature 84
5.2 The Base-emitter Voltage of the 5 V Design versus Temperature . . . 85
5.3 The Auxiliary Bandgap Voltage of the 5 V Design versus Temperature 85
5.4 The Reference Voltage of the 5 V Design for Various Loads at 27 ^{o}C . 86
5.5 The Transient Response of the 5 V Design at 27 ^{o}C (Input Rise Time of 1 μs) 86
5.6 The Response of the 5 V Design to a Supply Voltage with 0.1 V of Variations (at 27 ^{o}C) 87
5.7 The Response of the 5 V Design to a Supply Voltages of 5.1, 5.0 and 4.9 V 87
5.8 The Reference Voltage of the 3 V Design versus Temperature 88
5.9 The Base-emitter Voltage of the 3 V Design versus Temperature . . . 88
5.10 The Auxiliary Bandgap Voltage of the 3 V Design versus Temperature 89
5.11 The Reference Voltage of the 3 V Design for Various Loads at 27 ^{o}C . 89
5.12 The Transient Response of the 3 V Design at 27 ^{o}C (Input Rise Time of 1 μs) 90
5.13 The Response of the 3 V Design to a Supply Voltage with 0.1 V of Variations (at 27 ^{o}C) 90
5.14 The Response of the 3 V Design to a Supply Voltages of 3.1, 3.0 and 2.9 V 91

6.1 The Output Voltage of the 5 V Supply Bandgap Reference with Non-ESD-protected Pads 99
6.2 The Linearized Base-emitter Voltage of the 5 V Supply Bandgap Reference with Non-ESD-protected Pads 100
6.3 The Auxiliary Bandgap Voltage of the 5 V Supply Bandgap Reference with Non-ESD-protected Pads 101
6.4 The Output Voltage of the 5 V Supply Bandgap Reference with ESD-protected Pads 101

6.5 The Linearized Base-emitter Voltage of the 5 V Supply Bandgap Reference with ESD-protected Pads 102
6.6 The Auxiliary Bandgap Voltage of the 5 V Supply Bandgap Reference with ESD-protected Pads 102
6.7 The Output Voltage of the 3 V Supply Bandgap Reference with Non-ESD-protected Pads 103
6.8 The Linearized Base-emitter Voltage of the 3 V Supply Bandgap Reference with Non-ESD-protected Pads 103
6.9 The Auxiliary Bandgap Voltage of the 3 V Supply Bandgap Reference with Non-ESD-protected Pads 104
6.10 The Output Voltage of the 3 V Supply Bandgap Reference with ESD-protected Pads 104
6.11 The Linearized Base-emitter Voltage of the 3 V Supply Bandgap Reference with ESD-protected Pads 105
6.12 The Auxiliary Bandgap Voltage of the 3 V Supply Bandgap Reference with ESD-protected Pads 105

Chapter 1

Introduction

A voltage reference can be defined as an electronic circuit that produces a temperature-stable output voltage. An important feature of a voltage reference is its temperature-invariant property. However, the term *voltage reference* is often used with another commonly-used term in electronics, namely, voltage source which has a different emphasis. A voltage source focuses on providing an output voltage with an extremely low output impedance such that its voltage level is not affected by the type of loading. The design objective of a voltage source is to produce zero output impedance. In addition to voltage source, a voltage regulator is yet another term commonly confused with voltage reference. A voltage regulator is a circuit that provides a stable output voltage from a potentially noisy input supply voltage. It is normally used to power up other circuitry in a given application. Being so, a voltage regulator is required to have a high output current driving capability as opposed to voltage reference that cannot drive a heavy load. Given that the main objective of a voltage regulator is to provide power to other circuits, it does not have the same stringent requirement as does a voltage reference to provide a highly temperature-stable output voltage. In this thesis, the above distinctions between a voltage reference and the other two circuits is followed. When a voltage reference is mentioned, it is referred to an electronic

circuit that provides a highly temperature-stable voltage output.

Voltage reference is used in many electronic systems. It is needed whenever an application requires a certain reference for other voltage levels in a circuit to compare with. For instance, systems like voltage regulators, analog-to-digital converters, digital-to-analog converters, voltage-to-frequency converters, frequency-to-voltage converters, multimeters, transducer circuits, voltage-controlled oscillators, logarithmic amplifiers and other instrumentation and measurement circuits all require a reference voltage to function. In fact, the ultimate accuracy of these systems is limited by the precision of the voltage reference implemented within them. Therefore, the desire to have an accurate voltage reference is evident.

In order to characterize the performance of a voltage reference, a number of measurement parameters are used. These parameters help to describe the performance of a given reference. The commonly used parameters are line regulation, load regulation, temperature coefficient, ripple rejection ratio, long-term stability and output noise. Line regulation is defined as the ratio of the output voltage variation ΔV_o to the input voltage variation ΔV_i. It is a measure of a reference's ability to maintain a given voltage level under varying input conditions. Load regulation, on the other hand, is defined as the ratio of output voltage change ΔV_o to load current change ΔI_L. It is a means to measure the ability of a reference to sustain its output level over changing load condition. Temperature coefficient is an important parameter describing the performance of a voltage reference in terms of its capability to keep its voltage level over a given temperature range. It is defined as

$$TC(V_o) = \frac{\Delta V_o}{\Delta T} \tag{1.1}$$

in mV/°C. Yet, another commonly used form is

$$TC(V_o) = 10^6 \cdot \frac{\Delta V_o / V_o}{\Delta T} \tag{1.2}$$

in $ppm/°C$. The ripple rejection ratio tells a reference's ability to reject a ripple occurring in the supply voltage. It is a measure of how much ripple can pass through a reference from its supply to its output and is defined as

$$RRR = 20 \cdot log\left(\frac{V_{ri}}{V_{ro}}\right) \tag{1.3}$$

where V_{ri} and V_{ro} are the ripple magnitudes of the reference's input and output, respectively. Long-term stability is a measure of a reference's ability to maintain a given value over time. It has a similar form as temperature coefficient. The main difference is that ΔT is now replaced with ΔTime, which is the time change. Output noise is a measure which gives the fluctuation of the output voltage over a given range of frequency. All of these parameters help to describe a reference's performance and help a designer to choose an appropriate reference for a given application. In this thesis, the work focuses on improving the temperature coefficient of a voltage reference. This is the primary design objective of a voltage reference. The other parameters are also important but they are not considered in this thesis.

1.1 Voltage Reference Implementation

In order to implement a voltage reference, an electronic element such as a Zener diode can provide a well-defined voltage level. However, a Zener diode has difficulties in maintaining the voltage level over varying thermal environment because the behaviors of electronic elements are temperature-dependent in nature. Therefore, some effort must be put to eliminate the effects of temperature fluctuations in order to achieve

a temperature-stable voltage reference. There are two main approaches commonly used to remove or at least minimize the effect of temperature variations in voltage reference design. The first approach is called *temperature-regulated* while the second called *temperature-compensated.*

As the name implies, a temperature-regulated reference design is trying to maintain a constant thermal environment for a reference to operate [1, 2, 3]. In order to guarantee this environment, the substrate of the reference is regulated at a temperature above the maximum expected ambient temperature. It is done by a thermal feedback system which monitors the temperature of the substrate and uses a heat-generating device to maintain the temperature. The temperature of the substrate is usually sensed by exploiting the -2 mV/$^{\circ}C$ temperature coefficient of a base-emitter voltage drop [4]. The elevated temperature of the substrate is kept by means of an on-chip power dissipating transistor. The reason for holding the substrate temperature above the maximum operating ambient temperature is that only power dissipating devices are available in integrated circuits. Therefore, the only way to achieve a constant thermal environment is to raise the temperature above the reference's highest expected operating ambient temperature. Because of this fact, a lot of power is consumed just for keeping a constant thermal environment. In a power-critical application, this approach becomes unacceptable. This approach also decreases the lifetime of the reference as it always dissipates a lot of heat. Because of these, the temperature-compensated approach provides a viable alternative.

A temperature-compensated reference is achieved by using two temperature- changing voltage sources with opposite temperature coefficients to compensate the variations of each other. This compensation is usually done with a scale factor to provide the proper matching between the two sources. This scale factor is normally

implemented with a resistance ratio in a monolithic circuit. The resistors in a monolithic circuit provide excellent matching and tracking characteristics which make them the perfect choice to implement a scale factor of any value. Since the temperature-compensated approach does not operate at a higher temperature than required, it consumes less power in general than the temperature-regulated one. However, references implemented by this approach do not usually have the same performance as those by temperature-regulated. Even with the help of a scale factor, it is difficult to match two drift sources. The two commonly used approaches for implementing a temperature-compensated voltage reference are a Zener reference and a bandgap voltage reference. While both of them have the same difficulty in matching two drift sources perfectly, the bandgap reference has the advantage of requiring a lower supply voltage. In this thesis, it is discussed why a bandgap voltage reference has this matching problem and propose a new compensation scheme to improve the performance of a bandgap reference design.

1.2 Monolithic Circuit Technologies

The two design approaches for a voltage reference can only be implemented in monolithic circuit technology. For the temperature-regulated approach, it is because the constant thermal environment cannot be achieved without the substrate temperature control. This control mechanism can only be possible with a monolithic circuit in which all circuit elements are sitting in the same substrate. For the temperature-compensated approach, the reason is that the scale factor must be as temperature-independent as possible and this depends on the matching and tracking of resistors which implement the scale factor. Monolithic circuit technology has the advantages of good matching and tracking characteristics. Because of these reasons, voltage

reference designs are always implemented in monolithic circuit.

In monolithic circuit technology, the types of circuit elements can be put together are always limited by the compatibility of the fabrication processes required for each type of circuit elements. Especially for active devices, it is often the case that only one kind of active device can be implemented in a single fabrication technology. For this reason, monolithic circuit technology is usually classified by the active devices it can make. The commonly available technologies are categorized as bipolar, NMOS, CMOS, JFET, BiCMOS and GaAs. Within each category, there are many variations exist. The variations can be the minimum feature size, the speed of transistor, the types of passive elements achievable, etc. For monolithic circuits, the kinds of feasible passive elements are resistors and capacitors due to the limitation of fabrication processes. In general, no inductor can be possible in monolithic technology. For this reason, circuits designed for monolithic implementation must be accommodated for this deficiency. It is also true that some technologies are targeted for digital circuit implementations while some for analog. There are also technologies which allow for both. The BiCMOS technology that this thesis is based on circuit element models is one example. This technology is a 0.8 μm BiCMOS technology provided by Northern Telecom through Canadian Microelectronic Corporation [5]. The models provided with this BiCMOS technology are used for simulating the bandgap voltage reference designs of the proposed approach.

1.3 Thesis Content

The rest of this thesis is organized as follows. In chapter two, two types of temperature-compensated voltage references are reviewed. They are Zener voltage reference and bandgap voltage reference. The merits of each of them are discussed. The the-

ory of first-order compensation technique in bandgap reference is explained and the reason for their poorer performance compared to temperature-regulated reference is described. Some practical circuit implementations in bipolar technology are given. Voltage references implemented in other technologies such as NMOS and CMOS are also studied. After that, the curvature-compensation techniques for bandgap reference proposed in the literature are discussed. These are followed by a brief presentation of the new curvature-compensation technique.

Chapter three formulates the bipolar translinear principle which is applied in this thesis to design part of a circuit required by the proposed compensation technique. Some practical bipolar translinear circuits are given. Two circuits that are used in the implementation of the new curvature-compensation technique are designed based on the bipolar translinear principle. Then, the non-ideal behaviors of translinear circuits are discussed. The MOS translinear principle which is an extension of the bipolar translinear principle is also formulated in this chapter.

In chapter four, the proposed compensation technique is described in detail. The theoretical derivation of the proposed technique is given. The circuit topology designed based upon the proposed compensation technique is provided. After that, two circuits designed for two different power supply applications based on this topology are given.

A short description of the circuit simulator, Spectre, used in this thesis is given in chapter five. The simulation results of the two circuits designed in chapter four are presented in this chapter. These circuits are simulated with the model parameters given in the 0.8 μm BiCMOS process from Northern Telecom. The performance of each circuit is discussed.

Chapter six discusses the layout tools and some of the precaution steps a designer

will follow to minimize the effect of parasitics. Specifically, the layout issues relative to bandgap voltage reference are explored. After that, the simulation results of the two circuits based on layout back-annotation are given and discussed. These results include the effect of parasitics extracted from the circuit layout.

A conclusion of this thesis is provided in chapter seven. Some future works extendable from the proposed compensation technique are also discussed in this chapter.

Chapter 2

An Overview of Temperature-compensated Voltage References

2.1 Introduction

In this chapter, both Zener voltage reference and bandgap voltage reference designs are reviewed. It is followed by the theory of first-order compensated bandgap voltage reference. The error resulted from first-order compensation is discussed. Three first-order compensated bandgap references in bipolar technology proposed in the literature are studied. Voltage references implemented in other technologies are also discussed. The ideas for curvature-compensation found in the literature are reviewed. It is followed by the introduction of the new curvature-compensation technique.

The basic idea behind the design of a temperature-stable voltage reference is to compensate the negative temperature drift of one voltage drop by the positive temperature coefficient of another voltage drop. Then, with proper scaling of the second voltage drop by a temperature-independent scale factor, the temperature variations of the combined output will be cancelled out. This gives a voltage reference with a nominally zero temperature coefficient. In monolithic integrated circuit design, there are three well-known temperature drift sources which can be exploited in voltage

reference applications. The first voltage source is the base-emitter voltage V_{BE} of bipolar junction transistor which has a negative temperature coefficient, typical of about $-2\ mV/°C$ [4]. The second voltage source is the thermal voltage V_T which is linearly proportional to absolute temperature (PTAT). This voltage is obtained from the voltage difference between two base-emitter junctions, namely, ΔV_{BE}. The third voltage source is the reverse breakdown voltage V_Z of a Zener diode which also exhibits a positive temperature coefficient, typical of $+2.5\ mV/°C$. Based on these three sources, the two obvious choices one can have is either the combination of V_{BE} and V_Z or that of V_{BE} and V_T.

The scale factor can be implemented by the use of monolithic resistors. However, the temperature coefficient of easily available resistors is usually very high and demonstrates nonlinear temperature dependency. Therefore, it is generally not an adequate candidate for scale factor implementation. On the other hand, the ratio of two resistors made by the same materials in monolithic circuits exhibits excellent tracking over temperature, usually with a temperature coefficient of the order of a few $ppm/°C$. It follows that the ratio between two resistors, hereby referred to as the "resistor ratio" technique, is frequently used as a temperature-independent scale factor for voltage reference design.

2.2 Zener Voltage Reference

The approach exploiting the opposite-polarity drift of V_{BE} and V_Z to design a temperature -stable voltage reference is commonly called *Zener reference*. Fig. 2.1 shows a simple circuit design utilizing this approach. The Zener diode D_Z is supplied by a constant current I_0 and produces a reverse breakdown voltage V_Z with positive temperature coefficient. Traversing across the loop, one can obtain the following

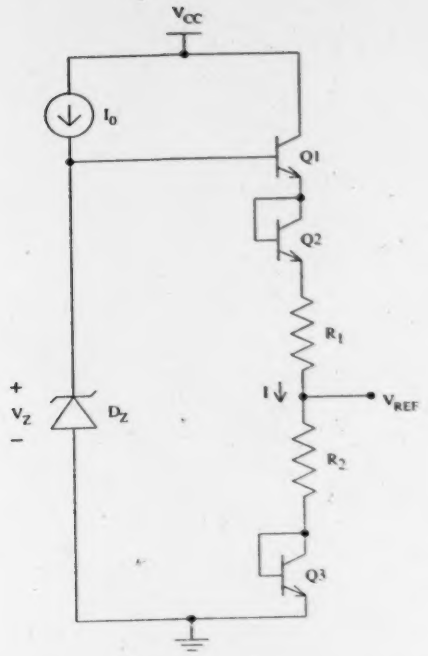

Figure 2.1: A Zener Voltage Reference Circuit

$$V_Z = V_{BE1} + V_{BE2} + I(R_1 + R_2) + V_{BE3} \tag{2.1}$$

Assuming all the transistors are matched, one can obtain

$$I = \frac{V_Z - 3V_{BE}}{R_1 + R_2} \tag{2.2}$$

Therefore, the reference voltage V_{REF} will be

$$V_{REF} = \frac{R_2}{R_1 + R_2} V_Z + \frac{R_1 - 2R_2}{R_1 + R_2} V_{BE} \tag{2.3}$$

This result shows that the temperature dependency of V_{BE} can be compensated by that of V_Z if proper resistor ratio is chosen. One can find the requirement for zero temperature coefficient by differentiating eqn.(2.3) with respect to temperature T and set the resulting equation to zero. The result is

$$\frac{R_1 - 2R_2}{R_2} = -\frac{\delta V_Z / \delta T}{\delta V_{BE} / \delta T} \tag{2.4}$$

Therefore, a voltage reference with nominally zero temperature drift can be achieved.

There are a number of drawbacks with using a Zener reference. First, since the breakdown voltage of a Zener is typically 7 V, it requires a power supply voltage of around 10 V to bring the Zener diode to its breakdown region. Therefore, a higher voltage supply is necessary. Second, a substantial amount of noise is introduced to the circuit by the Zener diode due to its avalanche breakdown operation. Finally, there is a long-term stability problem associated with the reference output voltage level due to the Zener diode. Thus, the absolute voltage level of the reference will change over time. Due to the introduction of the buried-Zener structures which have the diode operating well below the silicon surface, the long-term stability problem has been largely alleviated. The amount of noise is also reduced by the use of these structures. In spite of these improvements, the high supply voltage requirement still persists which often makes Zener references inapplicable in modern electronic circuits which frequently operate with a much lower supply voltage (usually 5 or 3 V). Therefore, another design approach which can operate with a lower supply voltage is essential for low voltage applications.

2.3 Bandgap Voltage Reference

The second approach for a temperature-stable voltage reference design is to compensate the negative temperature coefficient of V_{BE} by the thermal voltage V_T which has a positive coefficient. Voltage references of this kind are commonly referred to as a *bandgap voltage reference*. Fig. 2.2 gives a symbolic representation of a bandgap voltage reference. The base-emitter voltage V_{BE} with a negative temperature drift is established by the bias current I_1. Then, a positive drift due to the thermal voltage $V_T = kT/q$ is derived by the V_T generator. This voltage is further incremented K

times before it is added to V_{BE}. The output voltage V_{REF} is given as

$$V_{REF} = V_{BE} + KV_T \tag{2.5}$$

This equation illustrates that with an appropriate gain, K, the negative temperature coefficient of V_{BE} can be eliminated by the positive temperature coefficient of V_T.

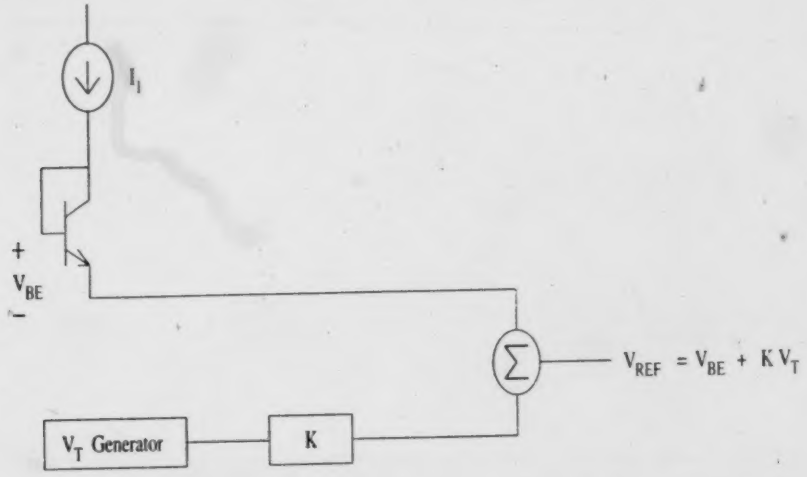

Figure 2.2: A Symbolic Representation of Bandgap Voltage Reference

The lower supply voltage requirement of the bandgap voltage reference can be seen through the following approximate analysis. The variation of V_T with respect to temperature can be found by differentiating the expression $V_T = kT/q$ and results in

$$\begin{aligned} TC(V_T) &= \frac{\delta V_T}{\delta T} \\ &= \frac{k}{q} \end{aligned} \tag{2.6}$$

By substituting $k = 1.38 \times 10^{-23}\ J/°K$ and $q = 1.60 \times 10^{-19}\ C$, the temperature coefficient of V_T is equal to

$$TC(V_T) = 8.625 \times 10^{-5}\ V/°K \tag{2.7}$$

The temperature coefficient of V_{BE} is approximately equal to [6]

$$TC(V_{BE}) = -2\ mV/°K \tag{2.8}$$

In order to compensate the negative coefficient of V_{BE}, the value of K in eqn.(2.5) must be equal to

$$\begin{aligned} K &= \frac{TC(V_{BE})}{TC(V_T)} \\ &= 23.188 \end{aligned} \tag{2.9}$$

At a temperature of 27 ^{o}C, the base-emitter voltage is approximately equal to 0.65 V and the thermal voltage is equal to 25.875 mV. Therefore, the reference voltage from eqn.(2.5) will be equal to

$$\begin{aligned} V_{REF(at\ 27\ ^{o}C)} &= V_{BE(at\ 27\ ^{o}C)} + KV_{T(at\ 27\ ^{o}C)} \\ &= 1.25\ V \end{aligned} \tag{2.10}$$

Since this voltage level is maintained throughout the temperature range of interest (except for the small error introduced by the nonlinearity explained later), the power supply voltage only need to have this value plus the headroom required for other circuitry in a bandgap reference circuit. Therefore, this approach is better for low power supply voltage applications.

In order to accurately compensate the drift, the value of K must be chosen in accordance to the negative drift of V_{BE}. Therefore, the temperature characteristics of V_{BE} must be studied more carefully. The base-emitter voltage of a bipolar transistor is related to its collector current by the equation [4]

$$I_C = CT^{\eta} \exp \frac{q(V_{BE} - V_{go})}{kT} \tag{2.11}$$

where C is a temperature-independent constant; T is the absolute temperature in ^{o}K; η is a fabrication technology dependent parameter related to doping level; and V_{go} is the bandgap voltage of silicon extrapolated to 0 ^{o}K.

If the bias current I_1 in Fig. 2.2 is a temperature-dependent current with the form

$$I_1 = AT^m \tag{2.12}$$

where A is a constant and m is the temperature dependency of I_1, then assuming negligible base current, one can substitute I_1 into I_C in eqn. (2.11). Then considering two temperatures: an arbitrary temperature T and a reference temperature T_r, one can easily derive the following equation

$$V_{BE}(T) = V_{BE}(T_r)\left(\frac{T}{T_r}\right) + V_{go}\left(1 - \frac{T}{T_r}\right) - (\eta - m)\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right) \tag{2.13}$$

where $V_{BE}(T_r)$ is the value of V_{BE} at the reference temperature T_r. If $V_{BE}(T)$ is written in the form of a constant term, a term proportional to T, and a higher-order term in such a way that the linear term represents the tangent to the $V_{BE}(T)$ at T_r, the following results

$$V_{BE}(T) = \left\{V_{go} + (\eta - m)\frac{kT_r}{q}\right\} - \lambda T + (\eta - m)\frac{k}{q}\left\{T - T_r - T\ln(\frac{T}{T_r})\right\} \tag{2.14}$$

where

$$\lambda = \frac{V_{go} + \frac{kT_r}{q}(\eta - m) - V_{BE}(T_r)}{T_r} \tag{2.15}$$

The second term in eqn.(2.14) is proportional to T which can be compensated by proper choice of K in eqn.(2.5). The third term is the nonlinearity and can be negligible in most applications. Fig. 2.3 shows the plot of this nonlinearity with $\eta = 3.54$ [4] and $T_r = 55\ ^\circ C$ ($328\ ^\circ K$) for various values of m in the temperature range $-55\ ^\circ C$ to $180\ ^\circ C$. As can be seen from the plot, the nonlinearity only introduces a few millivolts variations in the output voltage and is negligible in many applications. Therefore, after the compensation, the output voltage of the bandgap reference is

equal to the bandgap voltage V_{go} plus a constant term. That is why a reference done by this approach is called bandgap voltage reference.

In the design of a bandgap voltage reference, the values of various constants in eqn.(2.14) are unknown to the designer and vary for different technologies in use. These values are usually determined experimentally. Therefore, it would be more convenient to determine the value of K by observing the variation of V_{BE} to temperature through simulation or measurement.

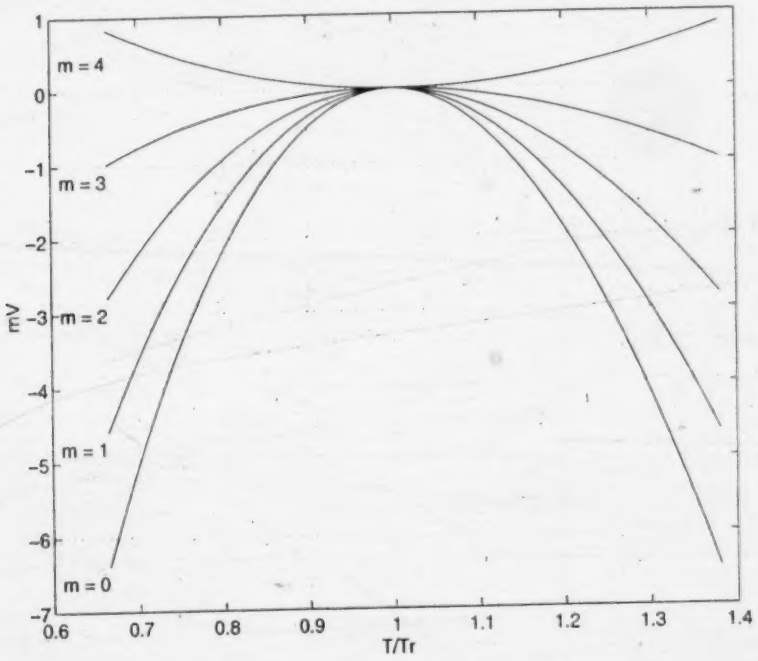

Figure 2.3: The Nonlinearity $(\eta - m)(k/q)\{T - T_r - T\ln(T/T_r)\}$ of $V_{BE}(T)$ vs. T/T_r

2.4 First-order Compensated Designs

In this section, some practical bandgap reference designs are discussed. The first design considered was proposed by Widlar [7] and is the first practical circuit that implemented the bandgap reference concept. The second circuit is proposed by Kuijk

[8] which eliminates some of the shortcomings in Widlar's design. The last one is the one by Brokaw [9] which is the most commonly used basic bandgap cell in many practical circuits (e.g. AD580, AD581, AD584, REF-01, AD573 and AD539). All these three designs are referred to as first-order compensated designs because they compensate for the linear temperature dependency of V_{BE} regardless of the nonlinear temperature variations. They are also the classical implementation of the bandgap reference concept that are described in many electronics textbooks.

2.4.1 Widlar Bandgap Reference

Fig. 2.4 shows a simplified form of the bandgap reference designed by Widlar [7]. This circuit utilizes a negative feedback by Q_3 to establish a nominal operating point such that the output voltage V_{REF} is equal to $V_{BE(on)}$ plus a voltage proportional to the difference between two base-emitter voltages. The operation of the feedback can be explained as follows. When the voltage level of V_2 is higher than the value in stable condition, more current will flow into the transistor Q_3 which causes less current going into the branch of R_2; thus reducing the current drive to the base of Q_3. V_2 will eventually decrease and restore to its stable value. On the other hand, if V_2 decreases, less current will flow into Q_3; thus more current will pass through R_2 and the voltage level of V_2 will be increased and eventually the nominal operating point is restored. When the nominal operating point is established, V_2 will be equal to $V_{BE(on)}$ of Q_3. At this point, Q_1 has a higher current density than Q_2 due to the lower value of R_1. This creates a voltage difference between the base-emitter voltages of Q_1 and Q_2 and this differential voltage ΔV_{BE} appears across the resistor R_3. If the transistors have high current gain β, the emitter current I_E of Q_2 will be almost

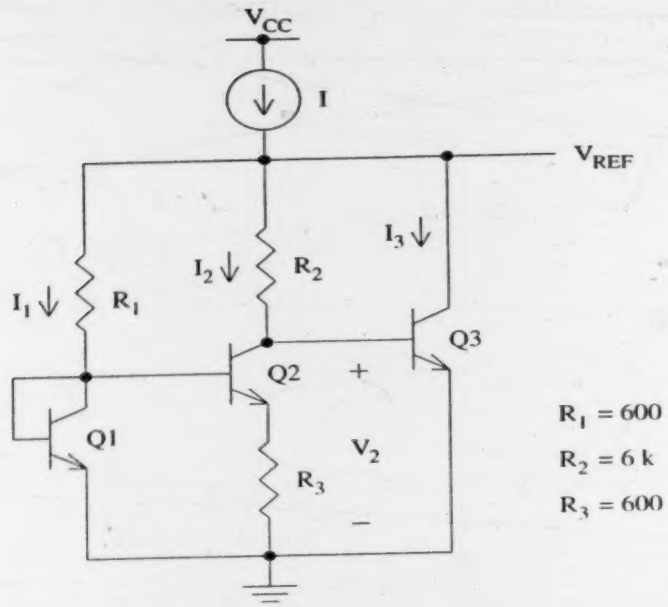

Figure 2.4: Widlar Bandgap Voltage Reference

equal to its collector current I_C. Thus,

$$\begin{aligned} I_2 R_3 &= V_{BE1} - V_{BE2} \\ &= V_T \ln \frac{I_1}{I_{S1}} - V_T \ln \frac{I_2}{I_{S2}} \\ &= V_T \ln \frac{I_1}{I_2} \end{aligned} \tag{2.16}$$

In eqn. (2.16), I_{S1} and I_{S2} are the reverse saturation currents of bipolar transistors Q_1 and Q_2. The reverse saturation current I_S of a bipolar transistor is equal to

$$I_S = ATn_i^2 \overline{\mu}_n \tag{2.17}$$

where

$$n_i^2 = BT^3 exp(-qV_{go}/kT) \tag{2.18}$$

and

$$\overline{\mu}_n = CT^{-n} \tag{2.19}$$

Since matched transistors are assumed, I_{S1} should be equal to I_{S2}. Therefore, the voltage across resistor R_2 will be equal to

$$\begin{aligned} V_{R2} &= I_2 R_2 \\ &= \frac{R_2}{R_3} V_T \ln \frac{I_1}{I_2} \end{aligned} \tag{2.20}$$

The output voltage V_{REF} is then

$$\begin{aligned} V_{REF} &= V_2 + V_{R2} \\ &= V_{BE(on)} + \frac{R_2}{R_3} V_T \ln \frac{I_1}{I_2} \end{aligned} \tag{2.21}$$

The above equation is in the form of eqn. (2.5) if the current ratio I_1/I_2 can be kept insensitive to temperature. Therefore, the first-order temperature dependence of V_{BE} can be reduced to zero by the second term in eqn. (2.21).

2.4.2 Kuijk Bandgap Reference

The bandgap design by Kuijk [8] is illustrated in Fig. 2.5. This circuit is operated with the feedback of the operational amplifier to establish an operating point. When the voltage V_+ at the positive input terminal of the op-amp is higher than the voltage V_- at the negative input terminal, the output voltage V_{REF} is increased which, in turn, triggers a higher current density through R_2. Thus, the voltage level at the negative terminal will be increased and finally equal to that at the positive terminal. At this point, a stable operating point is achieved. Alternatively, if V_- is higher than V_+, V_{REF} will go down. This will create a lower current density through R_2 and the voltage level of V_- will decrease which eventually stabilizes when $V_- = V_+$. When the stable operating point occurs, the voltages at the two inputs of the op-amp are equal. At this time, the voltage developed across resistor R_3 is equal to

$$\begin{aligned}\Delta V_{BE} &= V_{BE1} - V_{BE2} \\ &= V_T \ln \frac{I_1}{I_{S1}} - V_T \ln \frac{I_2}{I_{S2}} \\ &= V_T \ln \frac{I_1}{I_2} \frac{I_{S2}}{I_{S1}} \end{aligned} \tag{2.22}$$

Since the transistors are matched, I_{S1} and I_{S2} cancel out each other. So, the current through R_3 is equal to

$$I_2 = \frac{V_T}{R_3} \ln \frac{I_1}{I_2} \tag{2.23}$$

If the operational amplifier has high gain, $V_{in} \approx 0$ and $I_1 R_1 = I_2 R_2$. Then, the

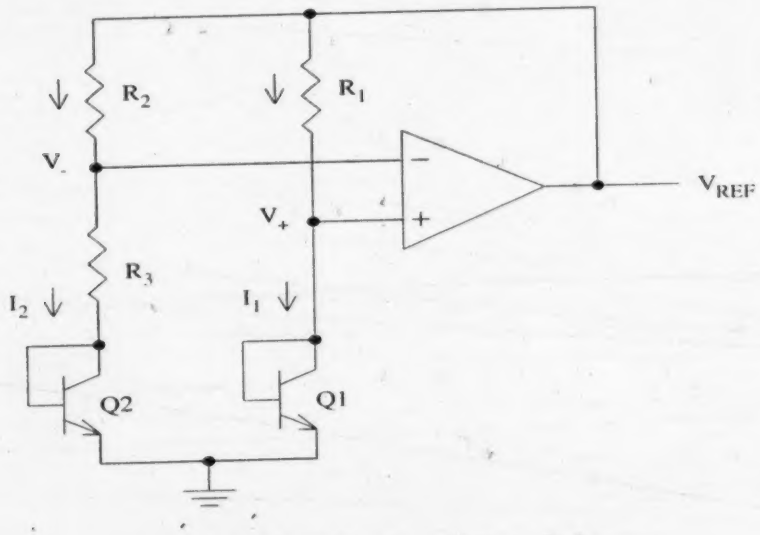

Figure 2.5: Kuijk Bandgap Voltage Reference

current ratio I_1/I_2 has the form

$$\frac{I_1}{I_2} = \frac{R_2}{R_1} \tag{2.24}$$

Using eqn. (2.23) and eqn. (2.24), one can easily obtain

$$V_{REF} = V_{BE1} + \frac{R_2}{R_3} V_T \ln \frac{R_2}{R_1} \tag{2.25}$$

In monolithic integrated circuit, the ratio of two resistances demonstrates excellent tracking over temperature. Therefore, the second term of eqn. (2.25) is linearly proportional to temperature such that the linear temperature variation of V_{BE1} can be compensated. Again, eqn. (2.25) is in the form of eqn. (2.5) with $K = R_2/R_3 \ln(R_2/R_1)$.

2.4.3 Brokaw Bandgap Reference

The third first-order bandgap design discussed is the one proposed by Brokaw [9], depicted in Fig. 2.6. In this circuit, the transistor Q_2 has an emitter area which is n times that of transistor Q_1. The circuit is operated as follows. When the voltage level at the bases of the two transistors is higher than what would be in quiescent point, a large current is forced through R_1. However, the voltage developed across R_2 will limit the current through Q_2 such that more current will flow through Q_1. The two different current densities in the collectors of Q_1 and Q_2 are sensed by the two resistors R. Therefore, a voltage imbalance is created at the two input terminals of the op-amp, and a lower voltage level appearing at the positive terminal which will then decrease the output voltage level of the op-amp; thus, restoring the quiescent point. When the base voltage is smaller than the quiescent value, a smaller current is forced through R_1. Now, Q_2 will take more current than Q_1 because of its larger emitter area. Again, the difference in current density of the two collectors is sensed by the two resistors R, causing a voltage imbalance in the inputs of the op-amp. The relatively lower voltage level at the negative terminal will cause the op-amp to increase its output drive to raise the base voltage, thus, restoring the quiescent operating point. Between these two conditions, the output of the reference has a stable output voltage which can be derived as follows.

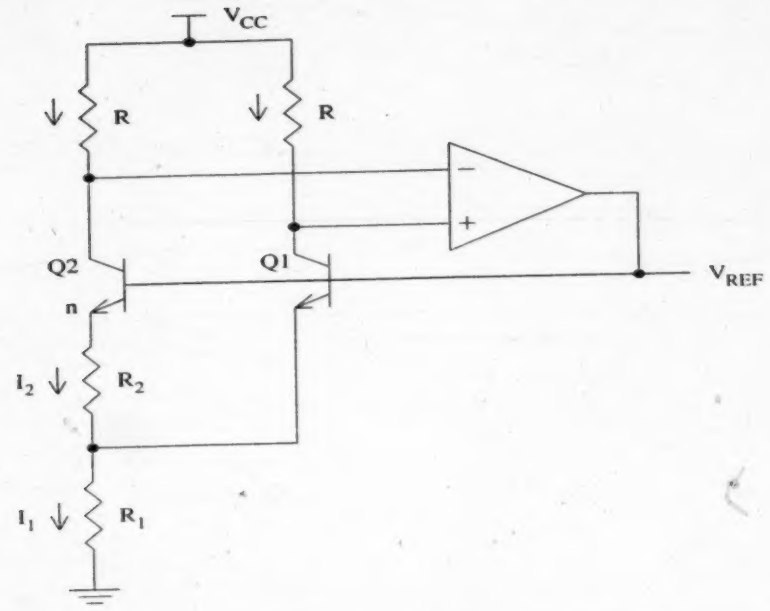

Figure 2.6: Brokaw Bandgap Voltage Reference

The voltage developed across R_2 is equal to

$$\begin{aligned} \Delta V_{BE} &= V_{BE1} - V_{BE2} \\ &= V_T \ln \frac{I_{C1}}{I_{S1}} - V_T \ln \frac{I_{C2}}{I_{S2}} \\ &= V_T \ln \frac{I_{C1}}{I_{C2}} \frac{I_{S2}}{I_{S1}} \end{aligned} \tag{2.26}$$

The saturation current ratio can be expressed in terms of the emitter area ratio which is equal to n. Since equal current is flowing in each transistor, $I_{C1} = I_{C2}$. Thus,

$$\Delta V_{BE} = V_T \ln n \tag{2.27}$$

Then, the current through R_2 is

$$I_2 = \frac{V_T}{R_2} \ln n \tag{2.28}$$

As the same current is flowing in Q_1, the current through R_1 is twice the current I_2

and the voltage across R_1 will be

$$\begin{aligned} V_{R1} &= 2\frac{R_1}{R_2}V_T \ln n \\ &= 2\frac{R_1}{R_2}V_T \ln n \end{aligned} \tag{2.29}$$

The ratio of two emitter areas in monolithic circuit is quite insensitive to temperature change. Thus, a voltage proportional to absolute temperature is established. The output of the bandgap reference will then equal to

$$\begin{aligned} V_{REF} &= V_{BE1} + V_{R1} \\ &= V_{BE1} + 2\frac{R_1}{R_2}V_T \ln n \end{aligned} \tag{2.30}$$

By proper selection of the constant coefficient of the second term in eqn. (2.30), a first-order temperature-compensated voltage reference is obtained.

2.4.4 Discussions

In the circuit designed by Widlar (Fig. 2.4), the base current is assumed to be negligible which is based on the assumption that the current gain β of transistor is high. However, this base current error can reduce the accuracy of the compensation. Since β varies with temperature, the base current error is different at various temperature and the analysis is not easy to include this effect. The second drawback of Widlar bandgap is that the operating currents in the circuit are all obtained from the current source I which is derived from the power supply. As the power supply fluctuates, this current will change. This will reduce the stability of the output voltage. The design by Kuijk eliminates the power supply problem by not generating the operating currents from the supply voltage. The current is derived from the base-emitter voltage differential and a resistor. This current is not affected by the supply voltage

variations as long as the transistors remain forward-active. In addition, the bias current to the transistors in Kuijk design is proportional to absolute temperature (eqn. (2.23) and eqn.(2.24)). The base-emitter voltage part of the reference is taken from Q_1 which is biased by the PTAT current. From Fig. 2.3, the nonlinearity error in V_{BE} is smaller when the bias current of a transistor is PTAT (m=1) than when it is independent to temperature. Therefore, the nonlinearity is smaller. However, the base-emitter voltage in Widlar's design is taken from Q_3 which is biased by a portion of I that is not a PTAT source. The bandgap reference proposed by Brokaw further eliminates the base current error by providing the current drive of the bases by the output of the op-amp. It also has the same advantage as that of Kuijk with a PTAT current as bias current to all transistors. The two latter circuits have an additional merit for being able to drive a certain amount of load by the op-amp. The Widlar reference, however, cannot be loaded and a buffer stage must be provided to connect this circuit to a load.

2.5 Other Technologies

The design approaches considered so far are implemented using bipolar technology. Since many analog circuits are designed using bipolar transistors, there will be no problem integrating a bandgap reference into other analog circuits on the same chip. However, the recent emphasis is in putting analog circuitry together with digital circuitry in one single chip; and cost-effective digital circuits are commonly implemented with CMOS technology. Hence, it is necessary to design analog circuit in CMOS transistors as well. For instance, in a data acquisition system, analog signal is frequently converted to digital domain by an analog-to-digital converter (ADC) and then processed by the digital circuits. Thus, it would be advantageous if the ADC

can be implemented by CMOS technology. This results in saving packaging cost of two chips and also circuit board area. In an ADC, there is always a requirement to have a stable reference voltage. This is one of the many reasons to design a voltage reference with only the devices available in CMOS technology.

In CMOS technology, there is a bipolar substrate transistor available with no modification to the basic process. This transistor is made by n-type substrate as collector, p-type well as base, and n^+ diffusion as emitter. Therefore, the topologies by Kuijk and Brokaw can be transferred to CMOS technology with only minor changes. As the collector of the transistor is made by the n-type substrate which is often tied to the most positive potential in a CMOS circuit, this device is restricted to a circuit configuration with its collector connected to the positive supply voltage. Therefore, a modified design is required. A typical circuit topology is shown in Fig. 2.7. As can be seen in the figure, the collector current can no longer be sensed directly as it does in the bipolar bandgap reference. It is instead sensed in the emitter. It follows that the error due to the finite current gain of transistor cannot be reduced. Thus, the accuracy of this bandgap circuit is reduced. However, this approach still provides an acceptable performance of bandgap reference in CMOS circuits.

While the bipolar configuration of bandgap reference can be readily implemented in CMOS technology, there are some other approaches proposed in the literature [10, 11, 12]. These approaches are invariably focused on generating the PTAT voltage from MOS transistors rather than using the base-emitter voltage differential ΔV_{BE}. In general, the PTAT voltage is obtained from the voltage difference between two NMOS transistors operating in weak inversion (or subthreshold region) with different bias current level.

When operated in weak inversion, the current in an NMOS is very small and the

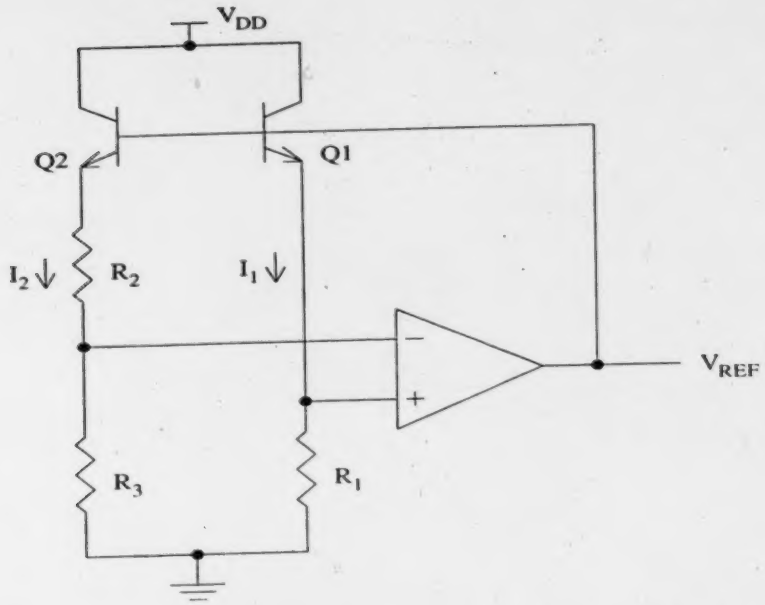

Figure 2.7: A Typical CMOS Bandgap Reference (with n-well process)

behavior of the transistor, with substrate connected to source, can be described by the following equation [10]

$$I_D = \left(\frac{W}{L}\right)\mu C_{ox}\left(\frac{1}{m}\right)\left(\frac{nkT}{q}\right)^2 \exp\left[\frac{q}{nkT}\left(V_{GS} - V_t - \frac{nkT}{q}\right)\right] \cdot \left[1 - \exp\left(\frac{-mqV_{DS}}{nkT}\right)\right] \tag{2.31}$$

where I_D is the drain current; V_{GS} is the gate-to-source voltage; V_{DS} is the drain-to-source voltage; W is the gate width; L is the gate length; C_{ox} is the oxide capacitance per unit area; μ is the effective mobility of carriers in the channel; V_t is the threshold voltage; k is the Boltzmann constant; q is the electron charge; T is the absolute temperature; m and n are process-related parameters. If a transistor is operated with $V_{DS} \gg kT/q$, the last exponential term in eqn.(2.31) disappears. When two NMOS transistors are operated in two different currents, with their gates connected

to the drain, the voltage difference between them is equal to

$$\begin{aligned} \Delta V_{GS} &= \left(\frac{nk}{q}\right) T \ln\left[\frac{I_{D1}}{I_{D2}} \cdot \frac{(W/L)_2}{(W/L)_1}\right] \\ &= AT \end{aligned} \tag{2.32}$$

This voltage has a positive temperature coefficient that can be exploited to compensate for the negative temperature coefficient of V_{BE}. It can also be scaled by choices of currents and device geometries. Fig. 2.8 depicts one implementation of this idea. The output voltage V_{REF} is equal to

$$\begin{aligned} V_{REF} &= V_{BE1} + V_1 - V_2 \\ &= V_{BE1} + AT \end{aligned} \tag{2.33}$$

Again, this voltage is in the form of eqn. (2.5).

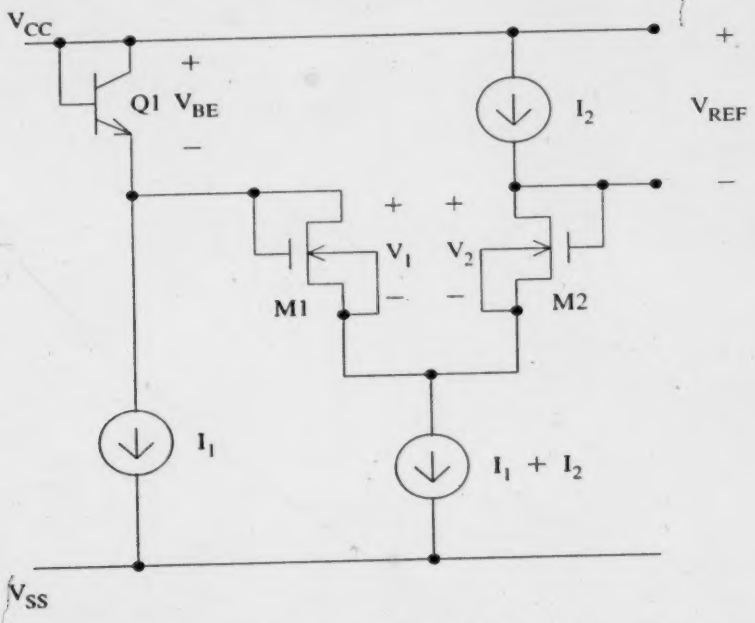

Figure 2.8: A CMOS Voltage Reference with MOS Transistors Operating in Weak Inversion

The application of this approach is only limited to reference operating in moderate temperature range. At high temperature, the junction leakage current of a MOS

transistor becomes comparable to, or even greater than, its current in weak inversion. Furthermore, the performance of voltage reference designed in this approach is not as good as those applying the direct bipolar bandgap design. These hinder the application of this design approach.

Voltage reference using NMOS technology has also been reported in the literature [13, 14]. The reference voltage is determined from the difference between the gate-to-source voltages of both the enhancement and depletion MOS transistors. A similar voltage drift versus temperature curve is obtained as a result of that the mobility variation dominates at low temperatures and the threshold voltage variation dominates at high temperatures. Therefore, the temperature coefficient is positive for low temperatures, returns to zero at the reference temperature, and then goes to negative for high temperatures. Voltage reference of this kind has a typical voltage level of 4 V and a temperature coefficient of $-50\ ppm/^{o}C$ [15]. The voltage level can be adjusted by changing the implant level that determines the threshold voltage.

In general, the design approaches discussed in this section are not as prominent as the bipolar bandgap. With the advent of BiCMOS technology where *npn* bipolar transistors are available, the same bipolar bandgap topology can be re-used in voltage reference design.

2.6 Curvature-Compensated Bandgap Reference

As Fig. 2.3 shows, the temperature variation of a theoretical bandgap reference voltage is in the range of a few millivolts. This error does not include the second order effects normally presented in physical components (like transistors and resistors) used in actual bandgap circuit. Therefore, a practical circuit will usually have an even higher error. When the precision of the reference voltage over temperature becomes

critical, there are a few measures that can be incorporated to reduce the error. These measures are focused on suppressing the influence of the following second order effects:

- The influence of the temperature-dependent base current (through β which is heavily sensitive to temperature).
- The Early effect.
- The effect of non-ideal resistors (which frequently has a certain degree of temperature dependency).
- The reverse Early effect (through the base-emitter junction)[19].

While reducing the second order effects can improve the accuracy of a bandgap reference, the overall precision is still limited to the few millivolts (shown in Fig. 2.3). This error is not due to second order effects occurred in the reference. Rather, it is a consequence of the basic bandgap reference design in which the nonlinear error term

$$(\eta - m)\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right) \tag{2.34}$$

is regarded small and is not compensated. Therefore, if a higher precision is desired, the nonlinear error must be compensated. In general, it is not easy to find a matching source to compensate this error in monolithic circuit. Compensation is then usually involved with certain degree of approximation.

The technique to compensate the nonlinear error term is commonly referred to as *Curvature-Corrected* or simply *Curvature Compensation*. There are a number of approaches proposed in the literature. These techniques can mainly be classified into two main groups. One involves finding an approximate nonlinear source with opposite temperature coefficient to compensate the error. Since this source is only an approximation, the overall error would not be completely eliminated but would be smaller. The other methods are focused on removing the nonlinearity in V_{BE} in the

first place. It is done by linearizing V_{BE} by a different circuit design.

A nonlinear term can be generated in a number of ways. One way is to use a MOS transistor [20] for the generation of the nonlinear term. Another way is to use a high TC resistor [21]. There are methods [22, 23] which only require the basic elements available in bipolar technology. The compensation technique described by Salminen and Halonen in [20] exploits the square-law $I_D - V_{DS}$ characteristics of the MOS transistor to compensate the nonlinear temperature variation of V_{BE}. A diode-connected MOS transistor is added to the basic bandgap cell and provides a nonlinear voltage drop for compensation. Audy [21], on the other hand, uses a high TC resistor connected in parallel with a low TC resistor to create two temperature-dependent voltage drops, one proportional to T^2 and one to T^3. These two voltage drops are used to compensate the nonlinearity in V_{BE}. The approaches that do not use special circuit elements are in general using two different currents to bias the two transistors which generate the voltage ΔV_{BE}. These two currents do not differ with each other by a constant but, instead are each varying differently with temperature. Thus, a term equal to $\frac{kT}{q} \ln\left(\frac{I_{C1}}{I_{C2}}\frac{I_{S2}}{I_{S1}}\right)$ is obtained. This term can be changed by varying the two currents as well as the size of the transistors. While changing the size of the transistors will alter the magnitude of this term, having a different temperature variation for the two bias currents will make this term nonlinear. Therefore, the nonlinearity in V_{BE} can be reduced by this term.

The methods that focus on linearizing V_{BE} are achieved by minimizing the effect of the nonlinear term in V_{BE}. This is accomplished by decreasing the magnitude of this term. As illustrated in Fig. 2.3, the error introduced by the nonlinear term is minimum if one can decrease the factor $(\eta - m)$. This term can be completely eliminated if one can have a current proportional to η. However, it is difficult if

possible to obtain this current in an actual circuit. In a monolithic circuit, a current proportional to $(\text{PTAT})^m$ (where m is an integer) can be generated by a special class of circuitry called *translinear circuit*. Therefore, a $(\text{PTAT})^m$ current with m close to η can be used to linearize V_{BE}. Again, the error is reduced.

In [24], a linearization method without resort to this $(\text{PTAT})^m$ current is proposed. It uses a PTAT current source to bias a series of m transistors and a temperature independent current source to bias another series of $m-1$ transistors. Then, the voltage differential between the two series of transistors are taken. The factor $(\eta - m)$ is obtained and can be reduced by proper choice of m. The number of transistors used is then dependent on the technology-related factor η. The drawback of this approach is that the supply voltage requirement will be larger when the number of transistors is increased.

In this thesis, the feasibility of using a proportional sum of two currents as bias current to a bipolar transistor is investigated. One of the currents is $(\text{PTAT})^m$ while the other $(\text{PTAT})^{m+1}$. These two currents are generated by a translinear circuit and are summed together in a proportion required by the technology parameter η. This approach would not only reduce the factor $(\eta - m)$ but also the argument of the logarithmic. The nonlinear error should be smaller than that using just one current source.

2.7 Conclusion

Both Zener reference and bandgap reference are studied in this chapter. In particular, the bandgap reference is considered extensively where the first-order compensation designs are reviewed. The implementations of first-order bandgap reference in different technologies are discussed. In addition, the curvature-corrected compensation

techniques of bandgap reference proposed in the literature are outlined. The main objective of these designs is to minimize the effect of the nonlinear term in the base-emitter voltage of a bipolar transistor. A new approach for curvature-corrected compensation to more effectively reduce the effect of the nonlinear term is introduced. This approach is explained in detail in chapter four.

Chapter 3

Translinear Circuits

3.1 Introduction

In this chapter, the translinear principle for bipolar transistors is formulated. Then, some typical circuit topologies are given. In particular, two circuits are designed based on the bipolar translinear principle. The two circuits provide polynomial functions for 5 V and 3 V supply voltage applications. They are utilized in the implementation of the new curvature compensation technique. Furthermore, the nonideal behaviors of TL circuits are discussed. Finally, the idea of extending the translinear principle to MOS transistors is provided.

The idea of translinear (TL) circuits was first proposed by Gilbert [25] in 1975. The term *translinear* stems from the relationship *"transconductance linear with current"*. According to Gilbert [26], a TL circuit is one that, with both inputs and outputs in the form of currents, exploits the linear relationship between transconductance and current in certain electronic devices, in an arrangement such that only these devices are presented in a closed loop, to achieve specific algebraic functions. One of these electronic devices possessing this relationship is the well-known bipolar junction transistor (BJT). As a consequence of the TL relationship, functions implemented by TL circuits are insensitive to temperature variations.

In addition to being temperature stable, bipolar TL circuits have another advantage of being superior in speed performance. It is a result of the small voltage swings in the circuit. Since a TL circuit is organized in loops of base-emitter junction voltages, the voltage swings are changes in the base-emitter voltages of transistors. The changes are due to alterations in collector (sometimes emitter) currents as signal currents vary and are usually a few tens of millivolts. Therefore, junction capacitances of transistors do not need to be charged and discharged significantly in performing circuit function. This results in high speed circuit performance, usually in a nanosecond range.

The applications of TL circuits are numerous. For instance, they can implement continuous-time algebraic functions like squaring; square-rooting; multiplication and division; multidimensional vector addition and subtraction; direct computation of amplitude ratios in an array; and polynomial, trigonometric and implicit-form function generation, to name a few. Although many of these functions can be implemented by a digital signal processing (DSP) approach, the TL circuit approach offers a high speed advantage. It also eliminates the conversion of signals from analog to digital domain before the function can operate on it; thus, removing an extra step to convert the digital signal back to analog form. In many applications, the information signals are not in digital form. In these cases, the TL circuit implementation provides an advantageous alternative for signal processing functions.

3.2 Bipolar Translinear Principle

The relationship between the base-emitter voltage (V_{BE}) of a BJT and its collector current (I_C) is described by the following equation

$$I_C = I_S \cdot \exp(\frac{V_{BE}}{V_T}) \tag{3.1}$$

and

$$I_S = A_E \cdot J_S(T) \tag{3.2}$$

where I_S is the saturation current of transistor; V_T is transistor thermal voltage; A_E is the emitter area of transistor and $J_S(T)$ is saturation current density. V_T is further given by

$$V_T = \frac{kT}{q} \tag{3.3}$$

where k is the Boltzmann's constant equals to 1.38×10^{-23} J/K (or 8.62×10^{-5} eV/K); q is the magnitude of an electronic charge and equals to 1.60×10^{-19} C and T is the absolute temperature in Kelvin, ^oK. Notice that eqn. (3.1) is the ubiquitous diode equation without the "-1" term. It is omitted here because transistors in a TL circuit are usually operated with the base-emitter junction forward-biased which implies $V_{BE} \gg V_T$. Therefore, the "-1" term becomes insignificant. If eqn. (3.1) is differentiated with respect to V_{BE}, it becomes

$$\frac{\delta I_C}{\delta V_{BE}} = \frac{I_C}{V_T} \tag{3.4}$$

Notice that eqn. (3.4) is the transconductance, g_m. Therefore, the g_m of a BJT is linearly related to the collector current as required by TL principle. Furthermore, if eqn. (3.1) is rewritten such that V_{BE} is in the left hand side, an expression for V_{BE}

is obtained

$$V_{BE} = V_T \cdot \ln \frac{I_C}{I_S} \tag{3.5}$$

Now, the bipolar TL principle can be formulated as follows. Fig. 3.1 demonstrates a generalized bipolar TL loop in which transistors are arranged in such a way that only the base-emitter voltages, $V_{BE}s$, are presented in the loop. For the principle to apply, the closed loop must have the following two conditions satisfied, namely

- the number of junctions has to be even.
- the number of clockwise-facing (CW) junctions must be equal to the number of anti-clockwise-facing (ACW) junctions.

The loop of Fig. 3.1 contains $2n$ base-emitter junctions and the junction voltages,

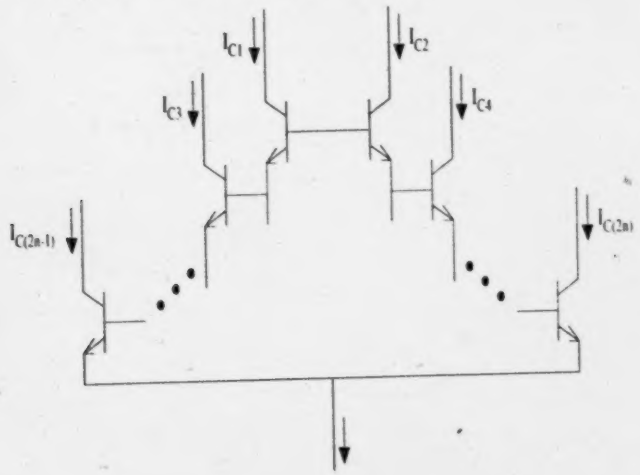

Figure 3.1: A Generalized Bipolar Translinear Loop

denoted here by V_{BEk}, must be summed to zero (by Kirchhoff's voltage law), that is

$$\sum_{k=1}^{2n} V_{BEk} = 0 \tag{3.6}$$

Here, V_{BEk}, is the base-emitter voltage given by eqn. (3.5). By substitution, eqn.

(3.6) becomes

$$\sum_{k=1}^{2n} V_T \cdot \ln \frac{I_{Ck}}{I_{Sk}} = 0 \tag{3.7}$$

The use of separate I_S for each junction allows that different emitter area can be used for different transistors. It also includes the possibility of using different device types (such as *npn* and *pnp* transistors), which have the same $V_{BE} - I_C$ relationship but different saturation current, in the same loop. When different types of devices are used, they must come in opposing pair, that is, one CW and one ACW. This ensures that the temperature dependent part of the saturation current $J_S(T)$ can be cancelled as it is different between two devices.

Now, eqn. (3.7) can be written in terms of the CW junctions and the ACW junctions

$$\begin{aligned} \sum_{CWk=1}^{n} V_T \ln \frac{I_{Ck}}{I_{Sk}} &= \sum_{ACWk=1}^{n} V_T \ln \frac{I_{Ck}}{I_{Sk}} \\ \prod_{CWk=1}^{n} \frac{I_{Ck}}{I_{Sk}} &= \prod_{ACWk=1}^{n} \frac{I_{Ck}}{I_{Sk}} \end{aligned} \tag{3.8}$$

In eqn. (3.8), the property that the summation of a series of logarithmic terms equals to the logarithmic of the product of terms is applied. Now, $I_{Sk} = A_{Ek} \cdot J_{Sk}$ can be substituted to eqn. (3.8) and, since J_{Sk} terms in the left hand side is equal to those in the right hand side, the following is achieved

$$\prod_{CWk=1}^{n} \frac{I_{Ck}}{A_{Ek}} = \prod_{ACWk=1}^{n} \frac{I_{Ck}}{A_{Ek}} \tag{3.9}$$

As obvious in eqn. (3.9), the two temperature-dependent terms, namely V_T and J_{Sk}, are not presented in the final expression. Therefore, the temperature sensitivity of this circuit is removed. The above equation is a compact form of the bipolar translinear principle. It states that: in a closed loop containing an even number of

base-emitter junctions where the number of clockwise-facing junctions is equal to the number of anti-clockwise-facing junctions, with no voltage generator inside the loop, the product of current densities in the clockwise-facing junctions is equal to those in the anti-clockwise-facing junctions. This principle is the backbone for analyzing a special type of circuits known as TL circuits which can perform complicated signal processing functions. Without it, the analysis of these circuits would be difficult.

3.3 Practical Circuits

In this section, some bipolar TL circuit topologies are discussed. The goal of this section is to demonstrate the feasibility of applying TL principle to implement some signal processing functions. In particular, two TL circuits that are used in this thesis are given. In the circuit diagrams followed, the symbol (+) simply means connecting this node to an appropriate potential. It does not mean to connect it to the power supply. In the examples shown, it is assumed that inputs to the circuits are available in the form of currents. No attempt is made to deal with the topic of generating the current signals. Unless otherwise stated, all the transistors considered are assumed to be the same size. It is further assumed that the base current of transistor introduces negligible error to the circuits.

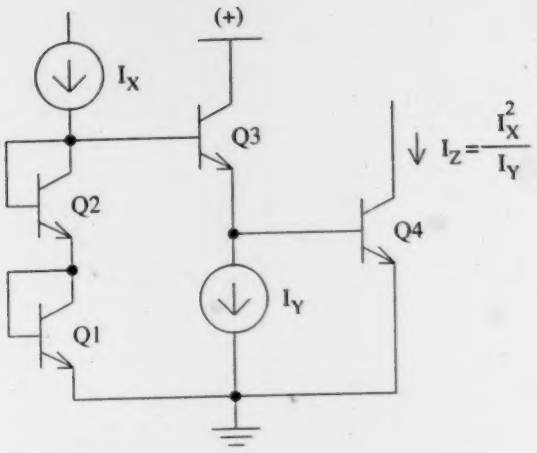

Figure 3.2: One-quadrant TL Squarer/Divider [26]

3.3.1 One-quadrant Squarer

Fig. 3.2 depicts a squaring circuit which can be analyzed as follows. Applying the TL principle, the product of currents in Q_1 and Q_2 is equal to that in Q_3 and Q_4. Therefore, the following is obtained

$$\begin{aligned} I_1 \cdot I_2 &= I_3 \cdot I_4 \\ I_X^2 &= I_Y \cdot I_Z \\ I_Z &= \frac{I_X^2}{I_Y} \end{aligned} \tag{3.10}$$

A squarer is thus obtained if I_Y is fixed. On the other hand, this circuit becomes a divider with I_Y as input if I_X is fixed. Therefore, this circuit is actually a squarer/divider. As evident in the schematic of this circuit, it is a one-quadrant squarer as the input currents must flow in the direction shown, otherwise, it will not function.

The circuit in Fig. 3.2 can be generalized to implement a function that raises the input signal to any degree, n, of power. By putting n diode-connected transistors in the branch where I_X is flowing and $n-1$ transistors with $n-2$ of them diode-

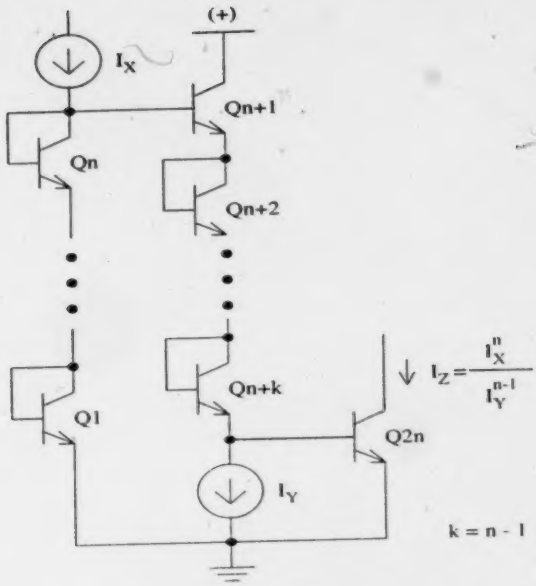

Figure 3.3: Power-n Circuit

connected in the I_Y flowing branch, a circuit which gives an output proportional to input signal raised to power n is obtained. Fig. 3.3 shows the schematic of this circuit.

Applying the TL principle, the following can be obtained

$$\begin{aligned} I_X^n &= I_Y^k \cdot I_Z \\ I_Z &= \frac{I_X^n}{I_Y^k} \\ &= \frac{I_X^n}{I_Y^{n-1}} \end{aligned} \tag{3.11}$$

Again, if I_Y is fixed, a power-n circuit is achieved.

This topology is very useful in obtaining a current proportional to any degree of a specific parameter when a current linearly dependent to the parameter is already available. For instance, if a linear temperature dependent current $I_1 = AT$ and a constant current $I_0 = B$ are presented, an output current having any degree, n, of temperature dependency can be generated. If I_1 is applied as I_X and I_0 as I_Y, the

following is obtained

$$\begin{aligned} I_Z &= \frac{(AT)^n}{B^{n-1}} \\ &= CT^n \end{aligned} \tag{3.12}$$

where $C = A^n/B^{n-1}$.

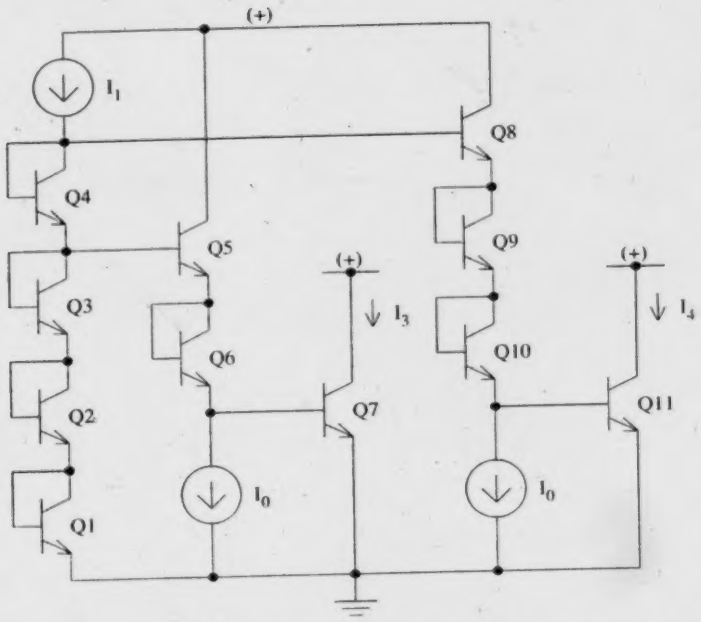

Figure 3.4: Circuit Generating Currents Proportional to T^3 and T^4 for 5 V Supply Voltage

In this thesis, currents proportional to T^3 and T^4 are needed. These currents, denoted here by, I_3 and I_4, can be realized easily by the topology in Fig. 3.3 when a current I_1 (proportional to T) and a constant current I_0 (independent of T) are given. In fact, the loops of TL circuits can be overlapped if there is a common loop between them. Since the I_1 branches in both circuits are common, they can be merged into one circuit providing the same currents. The circuit resulting from this combination is depicted in Fig. 3.4. This circuit is designed for supply voltage of 5 V. A second

circuit that provides the same function but for 3 V supply voltage is shown in Fig. 3.5. The generation of I_3 and I_4 in this circuit is achieved by first deriving the current I_2 (which is proportional to T^2). These two circuits are used in the implementation of the new compensation technique.

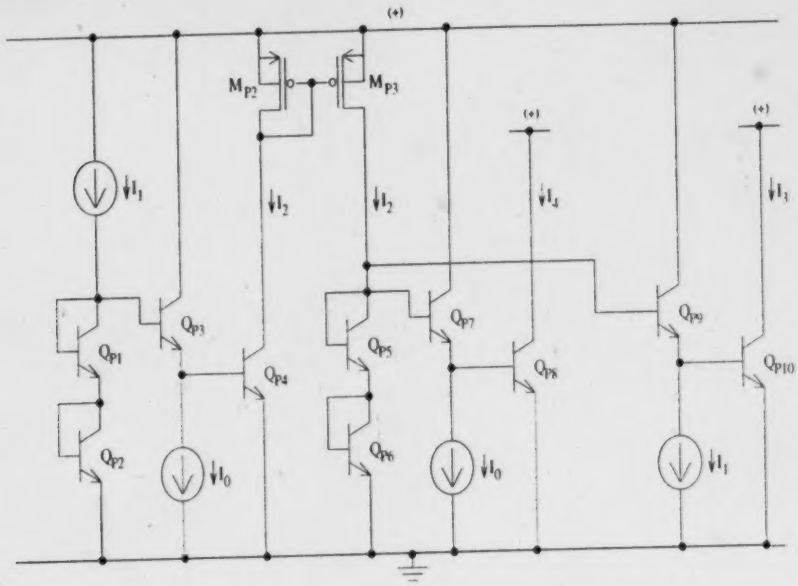

Figure 3.5: Circuit Generating Currents Proportional to T^3 and T^4 for 3 V Supply Voltage

3.3.2 Two-quadrant Squarer

The one-quadrant squarer can be made two-quadrant with an absolute-value circuit. One possible implementation of an absolute-value circuit is shown in Fig. 3.6(a). When the current I_X is flowing into the node, the current mirror reflects it to the output I_Z while transistor Q_3 is off. When I_X is flowing in the opposite direction, out of the node, Q_3 will be turn-on while the current mirror is off. Therefore, I_X is carried by Q_3 to the output. In both cases, I_Z flows in the direction shown. Therefore, the absolute-value function is achieved. V_{BIAS} is to ensure that Q_3 is off when I_X flows

into the node. It also has to ensure that Q_3 is on when I_X flows out of the node. In practice, the value of V_{BIAS} must be at least one V_{BE} above ground but it cannot be more than two V_{BE}s. Moreover, the output node must be at a potential level of at least one V_{BE} above ground so that Q_3 would not be saturated when I_3 of Q_3 is flowing out of the node. Therefore, the output of this circuit must be connected to an appropriate potential level. The circuit demonstrated in Fig. 3.6(b) reduces the finite-beta errors suffered from the circuit in Fig. 3.6(a) by introducing a Wilson current mirror and a Darlington pair.

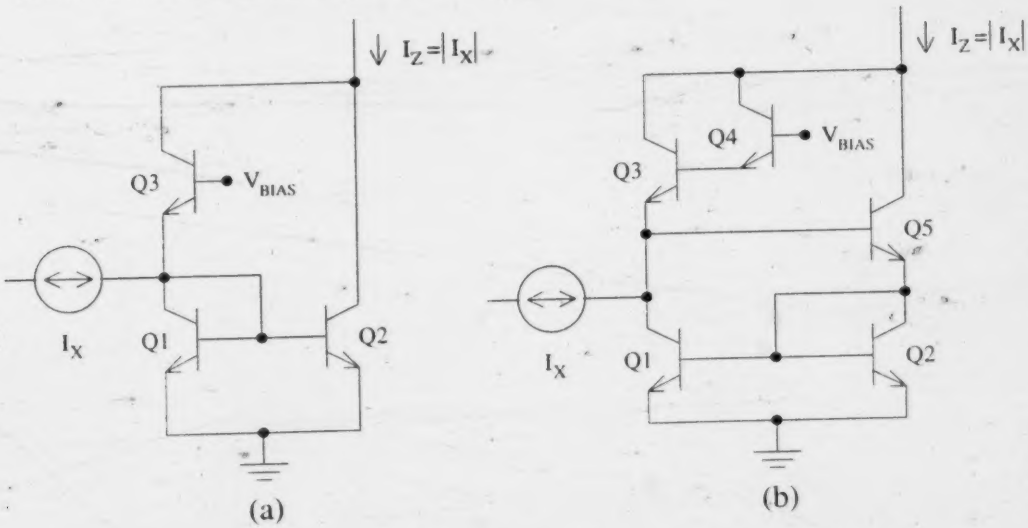

Figure 3.6: Absolute-value Circuits [26]

When the input current is a high speed signal of alternating polarity, the use of the absolute-value circuit to implement a two-quadrant squarer becomes unacceptable as transistors are needed to be switched off completely during one half cycle of the input. It will impair the speed of the circuit. The circuit in Fig. 3.7 provides a better approach of two-quadrant squarer for high speed application. It consists of two overlapping TL loops, $Q_1 - Q_2 - Q_3 - Q_7$ and $Q_6 - Q_5 - Q_4 - Q_7$. The input currents are two complementary signals, namely, $(1 + X)I$ and $(1 - X)I$ where

$-1 < X < +1$. Applying TL principle to the two loops,

$$\begin{aligned} I_{C3} \cdot I_{C7} &= I_{C1} \cdot I_{C2} \\ I_{C4} \cdot I_{C7} &= I_{C5} \cdot I_{C6} \end{aligned} \tag{3.13}$$

Let C be a temporary variable, so

$$\begin{aligned} I_{C3} &= (1+C) \cdot I \\ I_{C4} &= (1-C) \cdot I \end{aligned} \tag{3.14}$$

Solving eqn. (3.13) and eqn. (3.14),

$$I_{C7} = (1+X^2)I \tag{3.15}$$

Now, the output current I_Z will be equal to $(X^2)I$ if a current source I is applied as shown. In the example above. the input signal is in the form of a modulation factor. X. In general, this approach is quite often used for two-quadrant applications.

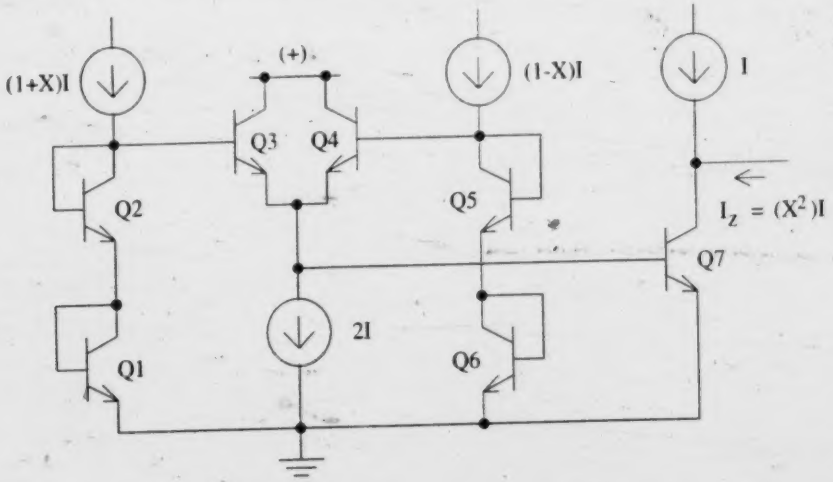

Figure 3.7: Two-quadrant TL Squarer [26]

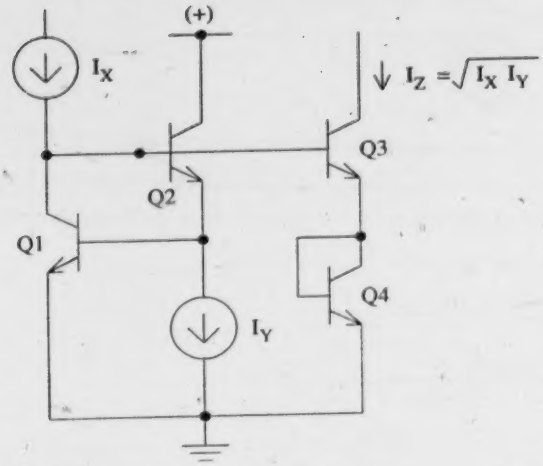

Figure 3.8: A TL Square-rooting and Geometric-mean Circuit [26]

3.3.3 Rooting and Geometric-mean Circuit

Fig. 3.8 illustrates a circuit implementing the rooting function. The output of the circuit is

$$I_Z = \sqrt{I_X I_Y} \tag{3.16}$$

If one of the input is fixed, this circuit implements the square-rooting function. However, when both inputs are allowed to vary, it will calculate the geometric-mean of the inputs. If the current I_X is proportional to T while the current I_Y is a constant current, the output current will be proportional to $T^{0.5}$. This implementation can be used only for a one-quadrant application which requires the input currents flows in the direction shown.

3.3.4 Vector Summation Circuit

Vector summation is another commonly used signal processing function. Fig. 3.9 depicts one implementation of this function by TL circuit. There are two TL loops

in this circuit. The loop involving I_X is described by the following

$$\begin{aligned} I_X^2 &= \frac{I_Z}{2} \cdot I_Z \\ I_X^2 &= \frac{I_Z^2}{2} \end{aligned} \tag{3.17}$$

while the loop containing I_Y can be described by

$$\begin{aligned} I_Y^2 &= \frac{I_Z}{2} \cdot I_Z \\ I_Y^2 &= \frac{I_Z^2}{2} \end{aligned} \tag{3.18}$$

Adding eqn. (3.17) and eqn: (3.18), one can derive

$$I_Z^2 = I_X^2 + I_Y^2 \tag{3.19}$$

Therefore, the output current I_Z is

$$I_Z = \sqrt{I_X^2 + I_Y^2} \tag{3.20}$$

This circuit has the advantage of being highly symmetrical which is desired in IC

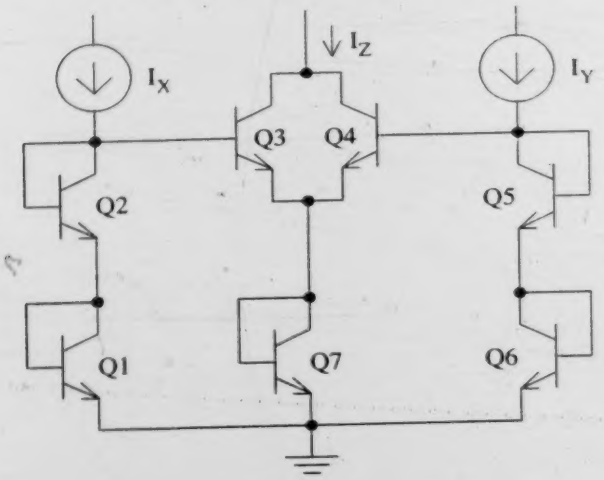

Figure 3.9: A TL Vector Summation Circuit [26]

operation. It is easily expandable to three or any dimensional vector summation.

Again, this circuit is applicable for one-quadrant input only. Therefore, the input currents must flow in the direction shown.

3.4 Non-ideal Behaviors

The variations of V_{BE} in a TL circuit cause the current ratio to be different from the theoretical derivation. This problem is commonly called "V_{BE} mismatch" [26]. It often happens as a result of errors in emitter area which are due to local variations in junction doping as well as errors in the photolithographic delineation of the emitter opening. The thermal gradients on a chip can also affect the value of V_{BE} since V_{BE} varies approximately $-2\ mV/°C$ [4]. The temperature variations in different part of a chip will make the V_{BE} values to be deviated from their designed ones and thus affecting the current ratio. Additionally, stresses in silicon can cause adjacent transistors to have different V_{BE} values. Although troublesome, these primary effects can be minimized by two measures. First, a highly symmetrical layout scheme can reduce errors in emitter area. Second, critical pairs of transistors should be arranged as cross-connected quads to lower the influence of thermal gradient on a chip.

Furthermore, V_{BE} is also affected by the collector-base voltage, V_{BC}, through base-width modulation (Early effect, named after J. M. Early who first interpreted this phenomenon). Likewise, a finite ohmic resistance in the base region (also called effective base resistance) increases V_{BE} to a higher value than that predicted by eqn. (3.5). These effects are collectively called secondary effects which are normally neglected in a first-order calculation. The circuits considered in the previous section do not include these effects.

The effect of V_{BC} is treated first. When the effect of V_{BC} is included in eqn. (3.4),

it becomes

$$V_{BE} = V_T \ln \frac{I_C}{I_S(T)(1 + \frac{V_{BC}}{V_{AF}})} \tag{3.21}$$

where V_{AF} is the forward Early voltage of transistor. As can be seen, the difference between eqn. (3.4) and eqn. (3.21) is the factor $(1+\frac{V_{BC}}{V_{AF}})$. When V_{AF} is very large, this factor approaches "1", making the effect of V_{BC} negligible. However, V_{AF} may be quite low (usually in the range of $5-50\ V$) in many high-frequency transistors. Therefore, even if transistors with $V_{AF} = 50\ V$ is used, a $1\ V$ error in the choice of collector bias will change I_C by 2%. Although the Early effect is very troublesome, it is not so predominant in TL circuits. Firstly, transistors in TL circuits are normally biased in pairs which effectively cancels the collector-biasing effects. Secondly, the collector is connected to the base in most transistors, thus making $V_{BC} = 0$, which essentially eliminates the Early effect. However, attention must still be paid to those transistors with collector not connected to the base. Finally, V_{AF} is in essence independent of temperature which further eliminates one possible chance of operating variability.

Now, the effect of base resistance, R_B is discussed. Taking R_B into account, the expression for V_{BE} becomes

$$V_{BE} = V_T \ln \frac{I_C}{I_S(T)} + \frac{I_C R_B}{\beta} \tag{3.22}$$

where β is the current gain, beta. Its value is a function of I_C, V_{CB}, as well as frequency. The effect of R_B is further complicated by the fact that the value of R_B is temperature dependent. As evident in eqn. (3.22), no longer can a closed-form equation for I_C (V_{BE}) be written. Therefore, errors introduced by R_B cannot easily be quantified in general terms as in the case of V_{BC}. Its effect should be considered under specific situations.

In addition, there is an error due to the finite beta of transistors. Since the base and collector of transistors in a TL circuit are often connected together with signal currents (inputs or outputs) driving the node, there is a small amount of current going to the base due to the finite beta of transistors. This loss of current from the collector will introduce a small error in the current ratio. In a high precision application, a base-current-cancellation technique should be incorporated to eliminate this error.

3.5 MOS Translinear Principle

Another electronic device that possesses a similar TL relationship, such as a BJT, is the well-known metal-oxide-silicon (MOS) transistor. The MOS transistors are commonly used in digital applications due to their smaller size and low power consumption. However, using MOS transistors for analog circuits is still not as common as using bipolar ones. It is mainly due to their poorer signal processing capability, a quality usually required by analog applications, compared to bipolar transistors. On the other hand, recent advance in providing bipolar transistors in a MOS process opens the possibility of using MOS transistors in the less crucial part of an analog circuit like biasing, while using bipolar transistors for the more critical signal processing part. In addition, improvements in performance of MOS devices may allow implementing signal processing functions by MOS transistors as well. Therefore, it would be advantageous if a similar TL principle can be extended to MOS transistors.

The first attempt to exploit the TL relationship of MOS transistors was due to Seevinck and Wiegerink [29]. Rather than having transconductance linear with current, MOS transistor has *"transconductance linear with voltage"* when it operates in strong inversion (saturation). This is evident in the square-law model for MOS

transistor operating in saturation. The square-law model gives

$$I_{DS} = \frac{\beta}{2}(V_{GS} - V_t)^2 \tag{3.23}$$

with

$$\beta = \mu C_{ox}\frac{W}{L} \tag{3.24}$$

where I_{DS} is the drain-to-source current; V_{GS} is the gate-to-source voltage; V_t is the device threshold voltage; μ is the effective surface mobility of the carriers in the channel; C_{ox} is the gate oxide capacitance; W and L are the width and length of device, respectively. Differentiating eqn. (3.23) with respect to V_{GS}, one gets

$$\frac{\delta I_{DS}}{\delta V_{GS}} = \frac{\beta}{4}(V_{GS} - V_t) \tag{3.25}$$

It shows that the transconductance, g_m, of MOS transistor is linearly related to the voltage $(V_{GS} - V_t)$.

This relationship was utilized by Seevinck and Wiegerisk [30] to formulate the MOS TL principle which can be derived as follows. A generalized MOS translinear loop is illustrated in Fig. 3.10. Although PMOS as well as a suitable combination of the both can be used in the loop, only NMOS transistors are used here for the sake of simplicity. In the loop, transistors are connected with their V_{GS} in series. In order to apply the TL principle, there are some requirements that must be met. They are the two conditions that must be satisfied in bipolar TL principle as well as all transistors in the loop must be in saturation. In summary, the conditions are:

- the number of gate-to-source junctions has to be even.
- the number of clockwise-facing junctions (CW) must be equal to the number of anti-clockwise-facing junctions (ACW).
- all transistors must be in saturation.

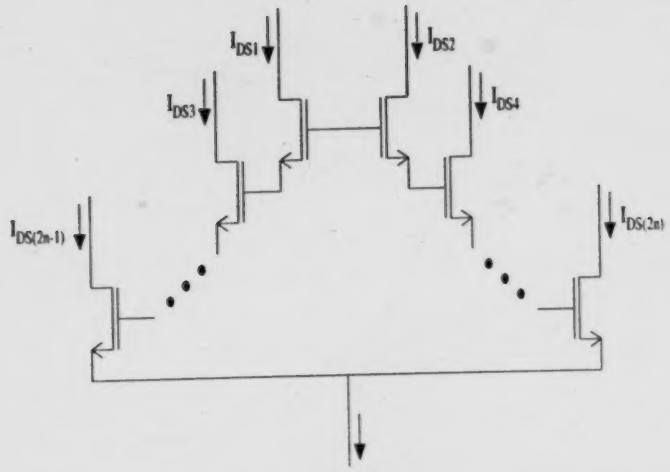

Figure 3.10: A Generalized MOS Translinear Loop

There are n gate-to-source junctions in CW direction as well as in ACW direction. Thus, applying Kirchhoff's voltage law to the loop, the following is obtained

$$\sum_{CWi=1}^{n} V_{GSi} = \sum_{ACWi=1}^{n} V_{GSi} \tag{3.26}$$

Again, the transistors are grouped into CW and ACW directions as in the bipolar case. Substituting eqn. (3.23) into eqn. (3.26) results in

$$\sum_{CWi=1}^{n} (V_{ti} + \sqrt{\frac{2 \cdot I_{DSi}}{\beta i}}) = \sum_{ACWi=1}^{n} (V_{ti} + \sqrt{\frac{2 \cdot I_{DSi}}{\beta i}}) \tag{3.27}$$

Since the same number of CW and ACW devices are presented in the loop, the threshold voltage terms on both sides of the equation can be cancelled. In monolithic circuit, well-matched threshold voltages are usually the case which justifies the above assumption. Furthermore, the parameters μ and C_{ox} for all devices are the same in a monolithic circuit. It follows that

$$\sum_{CWi=1}^{n} \sqrt{\frac{I_{DSi}}{(W/L)_i}} = \sum_{ACWi=1}^{n} \sqrt{\frac{I_{DSi}}{(W/L)_i}} \tag{3.28}$$

The above equation is a compact form of the MOS TL principle which can be exploited to implement a different set of nonlinear functions that the bipolar counterpart cannot

readily do. Notice that the MOS TL principle is a sum-of-roots relation rather than a product relation in the bipolar TL principle. Therefore, it does not lend itself to implement functions involving multiplication and division so easily. Since these two operations are fundamental to signal processing and many other functions, the utilization of MOS TL principle is less common. Furthermore, the performance of MOS circuits is inferior to their bipolar counterpart as is normally the case in analog design. In TL circuits, BJT can operate over a wider range of current while MOS transistors are bounded at the low end by weak inversion and at the high end by mobility reduction. Therefore, bipolar TL circuits have a wider dynamic range of operation than their MOS counterpart. It is also true that a tighter device matching can be achieved in bipolar devices. This is a fundamental requirement of TL principle. Due to these limitations, the use of MOS TL circuits has not been as successful as it is in bipolar circuits. On the other hand, MOS devices have the advantage of a zero DC gate current, thus, do not have the finite beta problem associated with bipolar TL circuits. In addition, more emphasis is placed in putting an analog circuitry in a digital circuit which is usually done by MOS devices. These may justify the use of MOS TL circuits rather than their BJT counterparts.

3.6 Conclusion

In this chapter, the bipolar translinear principle is formulated. Some typical signal processing circuits utilizing this principle are discussed. Especially, two TL circuits generating $(PTAT)^3$ and $(PTAT)^4$ currents are described. These circuits are utilized later in this thesis to implement the proposed compensation technique. Furthermore, the nonideal behaviors of the TL circuits are outlined. The possibility of extending the bipolar translinear principle to MOS devices is provided.

Chapter 4

A New Compensation Technique

4.1 Introduction

In chapter two, it is pointed out that the proposed curvature-compensation technique for bandgap reference is to use a proportional sum of two current sources, one $(\text{PTAT})^m$ and the other $(\text{PTAT})^{m+1}$, to bias a bipolar transistor (m is the closest smaller integer to the technology-related parameter η). These two currents are added in a proportion required by η. The base-emitter voltage V_{BE} obtained from this approach shows a better linearization over temperature. The linear portion of this voltage is compensated by a PTAT voltage such that the final voltage of the reference has only a small residual error.

The theoretical derivation of a new design approach is described in this chapter where $(\text{PTAT})^m$ and $(\text{PTAT})^{m+1}$ current sources are used to bias a transistor (with m being an integer). The condition required for eliminating the nonlinear error term in V_{BE} is determined. A special case is described that uses $m = 3$ to demonstrate the feasibility of the new method for linearization when η is equal to 3.54 [4]. The minimum error obtained with this approach is shown.

The derivation is followed by the design of circuit topology for implementing this new idea. Two similar circuit topologies are proposed with one designed for a power

supply voltage of 5 V while the other for 3 V supply voltage.

4.2 V_{BE} Linearization

As before, the collector current I_C of a bipolar transistor is related to its base-emitter voltage V_{BE} by the equation [4]

$$I_C = CT^{\eta} \exp \frac{q(V_{BE} - V_{go})}{kT} \tag{4.1}$$

where C is a temperature-independent constant; T is the absolute temperature in ^oK; η is a fabrication technology dependent parameter related to doping level; and V_{go} is the bandgap voltage of silicon extrapolated to 0 ^oK. If the bias current of the transistor is in the form of

$$I_C = AT^m + BT^{m+1} \tag{4.2}$$

where A and B are temperature-independent constants and m is an integer, then, by substitution, one can obtain

$$AT^m + BT^{m+1} = CT^{\eta} \exp \frac{q(V_{BE} - V_{go})}{kT} \tag{4.3}$$

By considering two temperatures: an arbitrary temperature T and a reference temperature T_r and substituting into eqn. (4.3), one can obtain two equations in terms of T and T_r, respectively. By dividing one equation by the other and solving for $V_{BE}(T)$, an equation for $V_{BE}(T)$ in the following form is achieved

$$\begin{aligned} V_{BE}(T) = & \left(\frac{T}{T_r}\right) V_{BE}(T_r) + V_{g0}\left(1 - \frac{T}{T_r}\right) \\ & + \frac{kT}{q}\left[ln\left(\frac{T^m + \rho T^{m+1}}{T_r^m + \rho T_r^{m+1}}\right) - \eta ln\left(\frac{T}{T_r}\right)\right] \end{aligned} \tag{4.4}$$

Here $\rho = B/A$. It is the proportion of the $(\text{PTAT})^{m+1}$ current to the $(\text{PTAT})^m$ current. Assume now that

$$T = T_r + \Delta T \tag{4.5}$$

where the ratio $\Delta T/T_r$ is small. Then eqn. (4.4) can be approximated using three terms of Taylor expansion for logarithms. This results in

$$V_{BE}(T) \approx V_{BE}(T_r) - \left\{V_{go} - V_{BE}(T_r) + \frac{kT_r}{q}\left[(\eta - m) - \frac{\rho T_r}{1+\rho T_r}\right]\right\}\left(\frac{\Delta T}{T_r}\right) - \left\{\frac{kT_r}{2q}\left[\eta - \frac{m + 2(m+1)\rho T_r + (m+1)(\rho T_r)^2}{(1+\rho T_r)^2}\right]\right\}\left(\frac{\Delta T}{T_r}\right)^2 \quad (4.6)$$

Hence, if the third term in (4.6) is zero then the temperature dependence of $V_{BE}(T)$ is linearized (with only residual third- and higher-order error terms). It is easy to find that this linearization condition requires that

$$\rho T_r = \frac{1}{\sqrt{m+1-\eta}} - 1 \quad (4.7)$$

This value of ρT_r will only be a positive real number if $m < \eta < m+1$. Therefore, the selection of m must meet this condition. However, there is no exact value for η as it varies with doping level; thus depending on technology. As a result, the choice of m as well as ρT_r are solely dependent on technology. Since the value of ρT_r is determined by η, the proportion of the two currents used for biasing transistor is also dictated by η.

In order to evaluate the achieved $V_{BE}(T)$ linearization, the condition in eqn. (4.7) can be substituted into eqn. (4.4) and found the residual error. To more effectively visualize the result, eqn. (4.4) is written in the form of a constant term, a PTAT term, and a higher-order term (H.O.T.) in such a way that the linear term represents the tangent to the curve of $V_{BE}(T)$ at T_r. The following equation is obtained

$$V_{BE}(T) = \left\{V_{go} + \frac{kT_r}{q}\left[(\eta - m) - \frac{\rho T_r}{1+\rho T_r}\right]\right\} - \lambda T + H.O.T. \quad (4.8)$$

where

$$\lambda = \left\{V_{go} - V_{BE}(T_r) + \frac{kT_r}{q}\left[(\eta - m) - \frac{\rho T_r}{1+\rho T_r}\right]\right\}\frac{1}{T_r} \quad (4.9)$$

and

$$H.O.T. = -\frac{kT_r}{q}\left[(\eta - m) - \frac{\rho T_r}{1+\rho T_r}\right] + \frac{kT_r}{q}\left[(\eta - m) - \frac{\rho T_r}{1+\rho T_r}\right]\left(\frac{T}{T_r}\right) - \frac{kT}{q}\left[(\eta - m)\ln\left(\frac{T}{T_r}\right) - \ln\frac{1+\rho T}{1+\rho T_r}\right] \quad (4.10)$$

The linear term in eqn. (4.8) can be compensated by a PTAT voltage which can be easily developed across a resistor. Therefore, the higher-order term given by eqn. (4.10) is the remaining residual error that cannot be cancelled by a PTAT voltage. In order to demonstrate the magnitude of this error, a special case that η equal to 3.54 is used. From eqn. (4.7), it is required that m equal to 3 for having a positive ρT_r. Therefore, the value of ρT_r is calculated. An expression of the residual error can be obtained by substituting these values into eqn. (4.10). This expression is plotted in Fig. 4.1 with $T_r = 55\ ^\circ C$ (328 $^\circ K$) in the temperature range of $-55\ ^\circ C$ to $180\ ^\circ C$.

As seen in this figure, the total residual error for the whole temperature range of interest is only 130 μV. This error is substantially lower than those of the first-order implementation which have an error of around a few millivolts (see chapter 2). Also evident in the figure is that the error is monotonically increasing with temperature. Therefore, there is a linearly increasing term in this error which can be absorbed by the compensating PTAT term. This linear term can be found by drawing a straight line between the two end points of this curve. When the compensating PTAT term is reduced by the same amount of this linear term, the final error that is achievable by this technique should be equal to the curve shown in Fig. 4.2. This is the theoretical result without considering the second order effects of the physical devices used in actual circuit implementation. The total residual error is now equal to 60 μV for the temperature ranging from -55 $^\circ C$ to 180 $^\circ C$.

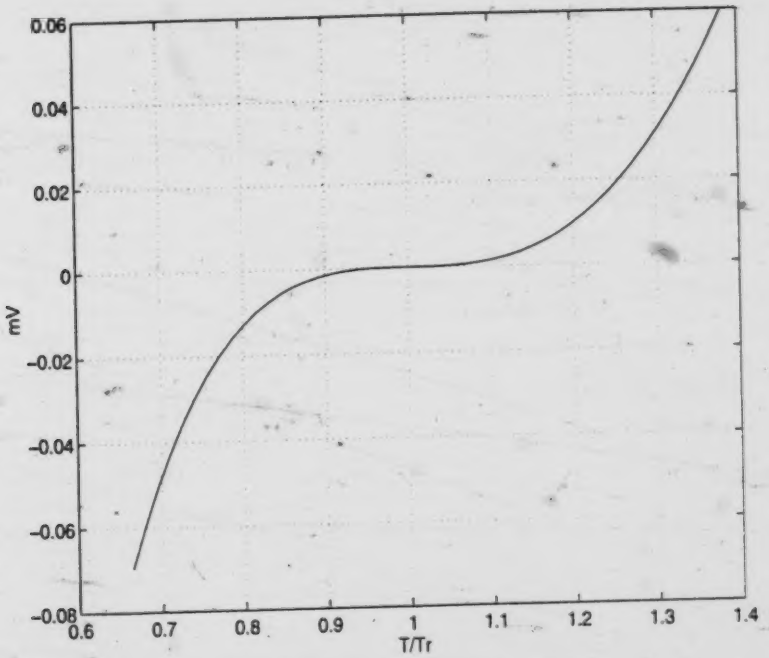

Figure 4.1: The Residual Error Resulted from the New Linearization Technique

Although the case that $\eta = 3.54$ is shown here, this technique is applicable to technology with any value of η. The conditions that have to be observed in any case are eqn. (4.7) and $m < \eta < m + 1$.

4.3 The Bandgap Reference Circuit Configuration

In this section, the basic configuration of the bandgap reference circuit for implementing the new compensation technique is presented. The main structure of the circuit in a simplified diagram format is first discussed. The functionality of each circuit block is explained briefly. After that, the exact implementations of the designs for both 5 V and 3 V supply applications are given. In this thesis, all the design circuits are implemented by Northern Telecom 0.8 μ BiCMOS process called Bipolar Analog Telecom Metal-Oxide-Semiconductor (BATMOS) technology. In this process, there

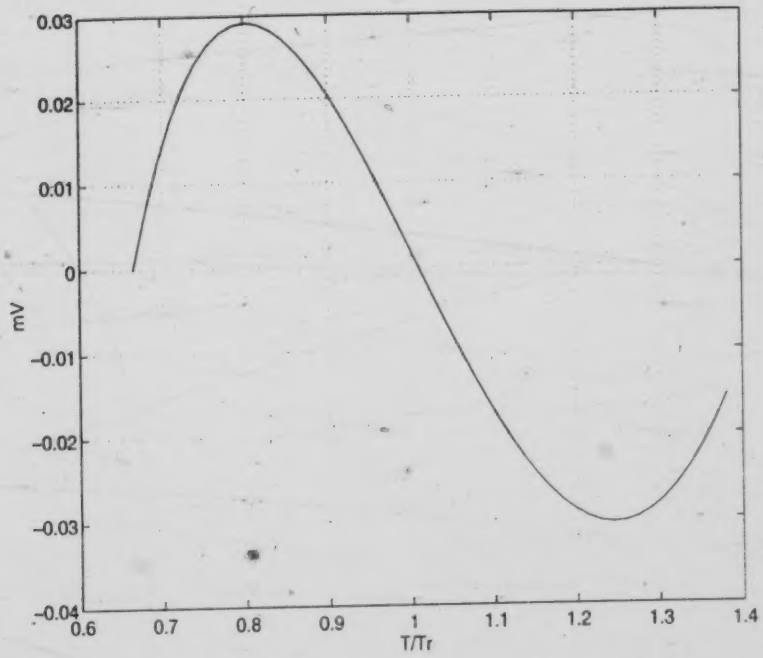

Figure 4.2: The Achievable Residual Error of the New Design Technique

is no pnp transistor available. It has only the npn bipolar transistors. Therefore, whenever an pnp transistor is needed, a PMOS transistor is often used to substitute it. The smallest feature size in this technology is 0.8 μm. However, the smallest feature size used in the new bandgap design is 2 μm. This is to conform the higher performance requirement normally needed in analog circuitry.

Fig. 4.3 shows the basic configuration of the bandgap circuit. It consists of an auxiliary bandgap reference circuit, a current summer, a translinear current polynomial circuit, a main bandgap circuit and some current mirrors. The auxiliary bandgap circuit is a Brokaw's bandgap reference. It consists of npn transistors (Q_{A1}, Q_{A2} and Q_{A3}), resistors (R_1, R_2 and R_3) and PMOS transistors (M_{A1} and M_{A2}). The letter M beside the transistor Q_{A1} denotes that it has an emitter area M times that of the others. The purpose of this circuit block is to produce two currents, a constant cur-

rent I_o and a PTAT current I_1 which are used by the translinear current polynomial circuit to produce two other currents, a $(\text{PTAT})^3$ current I_3 and a $(\text{PTAT})^4$ current I_4. The choice of currents I_3 and I_4 is implicitly implying that we have chosen m to be 3. This choice of m is justified by the fact that the value of η usually falls between 3 and 4 [4]. The operation of the auxiliary bandgap circuit is as follows. From the loop consisting Q_{A1}, Q_{A2} and R_1, the following equation is obtained

$$\begin{aligned} I_1 R_1 &= V_{BEA2} - V_{BEA1} \\ &= V_T \ln \frac{I_{A2}}{I_{SA2}} - V_T \ln \frac{I_{A1}}{I_{SA1}} \\ &= \frac{kT}{q} \ln \frac{I_{A2}}{I_{A1}} \frac{I_{SA1}}{I_{SA2}} \end{aligned} \tag{4.11}$$

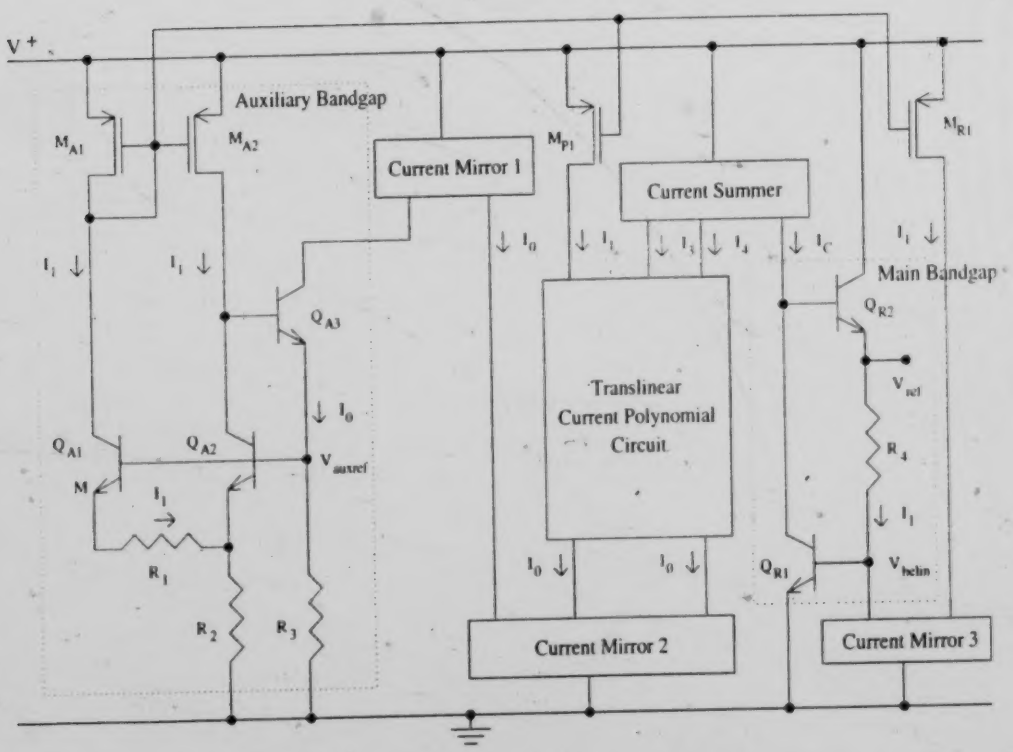

Figure 4.3: Bandgap Reference Circuit Configuration

Due to the action of the current mirror (M_{A1} and M_{A2}), the currents in both the

collectors of Q_{A1} and Q_{A2} are equal. Moreover, the current I_1 through R_1 should be equal to the collector currents of Q_{A1} and Q_{A2} provided that the base currents are negligible. Since the emitter area ratio of Q_{A1} and Q_{A2} is equal to M, the ratio of the saturation currents I_{SA1}/I_{SA2} should be equal to M as well. Therefore, a PTAT current is generated and equal to

$$I_1 = \frac{1}{R_1}\frac{kT}{q}\ln M \tag{4.12}$$

Now, two of these PTAT currents are flowing into R_2 and a PTAT voltage is created across R_2. This voltage is added to the base-emitter voltage V_{BEA2} of Q_{A2} to form the reference voltage V_{aurref} in the auxiliary bandgap circuit. Therefore, the auxiliary reference voltage is equal to

$$\begin{aligned} V_{aurref} &= V_{BEA2}(T) + 2I_1R_2 \\ &= V_{BEA2}(T) + 2\frac{R_2}{R_1}\frac{kT}{q}\ln M \end{aligned} \tag{4.13}$$

This voltage has an error of a few millivolts over the temperature range of interest and is considered negligible in this design. Therefore, a constant voltage is generated which is used to derive the constant current I_o by resistor R_3. Thus, the constant current is equal to

$$I_o = \frac{V_{aurref}}{R_3} \tag{4.14}$$

The transistor Q_{A3} is used to create a feedback such that the voltage V_{aurref} is stabilized. The currents I_o and I_1 are distributed throughout the circuit by current mirrors. These current mirrors are implemented by basic two-transistor current mirror circuit. Their exact configurations are provided for both 5 V and 3 V designs.

The constant current and the PTAT current are used by the translinear current polynomial circuit to generate both $(PTAT)^3$ and $(PTAT)^4$ currents. The implemen-

tation of this circuit is different for the 5 V and 3 V design due to the allowable supply voltage between rails in the circuit. The two different implementations are discussed fully later.

The main bandgap circuit consists of Q_{R1}, Q_{R2} and R_4. The transistor Q_{R1} is biased by the current I_C which is a proportional sum of I_3 and I_4 from the current summer. A better linearized base-emitter voltage is thus derived from V_{BER1}. This voltage is compensated by a PTAT voltage across R_4 which is obtained by forcing the PTAT current I_1 through the resistor. This PTAT current is brought in by current mirror 3. The transistor Q_{R2} is used to create a feedback in a fashion similar to that of Q_{A3} in the auxiliary bandgap circuit. The reference voltage V_{ref} is thus equal to

$$V_{ref} = V_{BER1} + \frac{R_4}{R_1}\frac{kT}{q}\ln M \qquad (4.15)$$

This voltage is the output of the new bandgap reference design.

4.3.1 5 V Supply Bandgap Reference Circuit

The complete implementation detail of the 5 V supply bandgap voltage reference is provided here. The various values of resistors in the circuit are derived. In addition, the sizes of various transistors in the circuit are determined and the reasons for their sizing are described. The complete schematic is shown in Fig. 4.4 where the labeling of transistors and resistors follows those in Fig. 4.3 in order to give a close correspondence between the two circuits. The discussion is started with the design of the auxiliary bandgap. Compared to that of Fig. 4.3, it is noticeable that there are some differences between the two figures. Firstly, transistor Q_{A1} is replaced by four transistors. As discussed previously, Q_{A1} is different from the other transistors by having an emitter area of M times that of the others. For a transistor with M times emitter area, it is equivalent to M transistors of the basic unit connected in parallel.

This is the reason for having 4 parallel transistors replacing Q_{A1}. The reason for M = 4 would become clear when the design parameters of the auxiliary bandgap are discussed. Secondly, there is a difference in Q_{A3}. It is replaced by two transistors connected in a Darlington configuration. This is used to increase the current drive in the output V_{auxref}. Finally, there are some additional transistors presented in the auxiliary bandgap of Fig. 4.4, namely, Q_{A4}, and the modified Wilson current mirror implemented by M_{A3}, M_{A4}, M_{A5} and M_{A6}. These transistors are used for the compensation of the base currents in Q_{A1} and Q_{A2}. The operation of these transistors is as follows. Q_{A4} senses the base current of Q_{A1} and this current is doubled by the current mirror which, in turn, feeds the current into the bases of Q_{A1} and Q_{A2}. The current doubling is achieved by making the width of M_{A4} and M_{A6} twice that of M_{A3} and M_{A5}. Here, both M_{A3} and M_{A5} are arbitrarily chosen to have a width to length ratio of 4:2. Therefore, the width to length ratio of both M_{A4} and M_{A6} must be 8:2 in order to double the current. While a larger value of length or width can be used, the above values are chosen because they require less silicon area while still suffice the requirement. However, a different ratio can also be used.

Now,the design of resistance values in all resistors are discussed. The choice of these resistances determine the parameters I_1, V_{auxref} and I_o in the auxiliary bandgap. As shown in eqn. (4.12), R_1 sets the values of I_1 at various temperatures if the value of M is already selected. Therefore, there are actually two parameters that can be selected to design the desired value of I_1 at various temperatures. There are many possible choices for the values of R_1 and M. The choice of R_1 is determined by the feasible size of the resistor since it determines the silicon area needed for implementation. At the same time, the choice of M is also dictated by the required area necessary to layout the desired number of transistors. Therefore, there is a trade-

off between these two parameters to achieve the desired value of I_1. In this design, the value of M is chosen to be 4 and R_1 to be 724 Ω. These values give an I_1 of 49.7 μA at 27 $^{\circ}C$. Since I_1 is a PTAT current, its value varies linearly with absolute temperature. For the temperature range of interest, I_1 has its lowest value of 36.9 μA at - 50 $^{\circ}C$ and highest value of 75.0 μA at 180 $^{\circ}C$. Thus, a reasonable value of I_1 is obtained with feasible values of M and R_1.

After setting the value for I_1, the second step is to design the value of V_{aurref}. The criterion is to design the second term in eqn. (4.13) such that the linear term of $V_{BEA2}(T)$ is compensated. As transistor Q_{2A} is biased by the current I_1 which has a linear temperature dependency, $V_{BEA2}(T)$ should have the form of eqn. (2.22) with $m = 1$. Thus, the following is obtained

$$V_{BEA2}(T) = \left\{V_{go} + (\eta - 1)\frac{kT_r}{q}\right\} - \lambda T + (\eta - 1)\frac{k}{q}\left\{T - T_r - T\ln(\frac{T}{T_r})\right\} \quad (4.16)$$

where

$$\lambda = \left\{V_{go} + \frac{kT_r}{q}(\eta - 1) - V_{BEA2}(T_r)\right\}\frac{1}{T_r} \quad (4.17)$$

In order to compensate the linear term, the second term in eqn. (4.13) plus the second term in eqn. (4.16) should be equal to zero. Therefore, the following condition has to be met

$$2\frac{R_2}{R_1}\frac{k}{q}\ln M = \left\{V_{go} + \frac{kT_r}{q}(\eta - 1) - V_{BEA2}(T_r)\right\}\frac{1}{T_r} \quad (4.18)$$

Since the values for R_1 and M are already selected, there is only R_2 left for adjustment. As shown in eqn. (4.18), the adjustment requires knowledge of V_{go}, $V_{BE}(T_r)$ and η. The value of $V_{BE}(T_r)$ can be obtained either through simulation or experimental measurement on the transistor by biasing it with the current I_1 and observe its base-emitter voltage at the reference temperature T_r. However, the values of V_{go} and η

require an empirical measurement of the target device and are usually not available to circuit designer. However, some commonly referenced values like $V_{go} = 1171$ mV and $\eta = 3.54$ [4] can be used as the initial design parameters to set R_2. The value of R_2 can then be varied during simulation to achieve the desired reference temperature. In this design, the value of R_2 is found to be 4 $k\Omega$.

Now comes to deciding the value of R_3 to set the desired I_o. As given in eqn. (2.28), a first-order bandgap reference usually has a voltage level of around 1.25 V. This value can be assumed as the initial design of I_o. The choice of I_o is determined by the current level that can be sustained by Q_{A3} without saturation. At the same time, this current is used by the translinear circuit to generate I_3 and I_4. If it is too large, the current level of I_3 and I_4 will become too low to be feasible in practice. In this design, it is chosen for $R_3 = 11.3$ $k\Omega$ to obtain an I_o of 110.6 μA. To finish the design of the auxiliary bandgap reference, it is needed to choose the sizes of the PMOS transistors in the two current mirrors implemented by M_{A1} and M_{A2}, and M_{M1} and M_{M2}, respectively. The sizes of transistors in the two current mirrors must be chosen in such a way that the PMOS transistors operate in saturation region. This ensures that the currents flowing through them do not change with V_{DS}. The equation describing a PMOS transistor operating in saturation is given as

$$I_{DS} = \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{GS} - V_t)^2 \qquad (4.19)$$

with the condition that $(V_{GS} - V_t) < V_{DS}$. Since I_{DS} is the current set by I_1 or I_o in either case, a ratio of W:L is selected in a way which ensures the condition to be met at all time. In this design, W:L is chosen to be 40:2 for all the four transistors to ensure that all of them are in saturation.

Now, the translinear current polynomial circuit design is described. Since the

main objective of this circuit is to provide the two currents proportional to T^3 (I_3) and T^4 (I_4) from the currents I_o and I_1, the circuit from Fig. 3.4 can be used. This circuit accepts I_o and I_1 to provide I_3 and I_4. Therefore, the main part of this circuit used here is the same as the one given in Fig. 3.4. The current I_1 is brought in by M_{P1} while the current I_o by the current repeater Q_{M21}, Q_{M22} and Q_{M23}. The generation of the two currents can be explained as follows. There are two translinear loop in this circuit. One is formed by Q_{P1}, Q_{P2}, Q_{P3}, Q_{P5}, Q_{P6} and Q_{P7}. Applying the TL principle, the equation is obtained

$$I_{C1} \cdot I_{C2} \cdot I_{C3} = I_{C5} \cdot I_{C6} \cdot I_{C7} \tag{4.20}$$

Since I_1 is flowing in Q_{P1}, Q_{P2}, and Q_{P3} and I_o flowing in Q_{P5} and Q_{P6}. eqn. (4.20) becomes

$$I_1^3 = I_o^2 \cdot I_3 \tag{4.21}$$

where I_{C7} is replaced by I_3. So, the current I_3 becomes

$$I_3 = \frac{I_1^3}{I_o^2} \tag{4.22}$$

Substituting eqn. (4.12) and eqn. (4.14), I_3 is equal to

$$I_3 = \left(\frac{k}{qR_1}\ln M\right)^3 T^3 \left(\frac{R_3}{V_{aurref}}\right)^2 \tag{4.23}$$

The other translinear loop is formed by Q_{P1}, Q_{P2}, Q_{P3}, Q_{P4}, Q_{P8}, Q_{P9}, Q_{P10} and Q_{P11}. Applying a similar analysis as for I_3, the following is obtained

$$I_4 = \left(\frac{k}{qR_1}\ln M\right)^4 T^4 \left(\frac{R_3}{V_{aurref}}\right)^3 \tag{4.24}$$

These two currents are scaled and added together by the current summer (M_{S1}, M_{S2}, M_{S3} and M_{S4}). The sizes of M_{S1} and M_{S3} remain the same as before, that is, are

equal to a length to width ratio of 40:2. The lengths of M_{S2} and M_{S4} are both equal to 2 μm while the widths are different. This is to provide a proper scaling factor for I_3 and I_4 such that these two currents are in a suitable proportion required by η for optimal linearization of V_{BE} in the main bandgap. The width of M_{S2} is chosen to 40 μm while that of M_{S4} is 130 μm. These values are determined from simulation result and may be varied for different technology. The two currents are then added together and denoted by I_C in Fig. 4.4. This is the bias current of Q_{R1} and is given by

$$I_C = AT^3 + BT^4 \tag{4.25}$$

where

$$A = \left(\frac{k}{qR_1}\ln M\right)^3 \left(\frac{R_3}{V_{auxref}}\right)^2 \tag{4.26}$$

and

$$B = \frac{130}{40}\left(\frac{k}{qR_1}\ln M\right)^4 \left(\frac{R_3}{V_{auxref}}\right)^3 \tag{4.27}$$

By this current, the base-emitter voltage V_{BER1} of Q_{R1} is linearized. The linear temperature dependency of V_{BER1} can then be compensated by a PTAT source to achieve a constant reference voltage. These are done in the main bandgap circuit. The value of ρ in this case can be found by dividing eqn. (4.26) by eqn. (4.27) and is equal to

$$\rho = \frac{130}{40}\left(\frac{k}{qR_1}\ln M\right)\left(\frac{R_3}{V_{auxref}}\right) \tag{4.28}$$

As obvious in eqn. (4.28), the value of ρ would be practically temperature independent if R_1 and R_3 are tracking each other.

The main bandgap circuit consists of the same circuit elements as those in Fig. 4.3 with some additional elements to improve the performance of the bandgap reference.

Transistor Q_{R3}, M_{R3} and M_{R4} are used to compensate the base current of Q_{R1}. Both M_{R3} and M_{R4} have the same width to length ratio of 8:2. Transistor Q_{R2} is replaced by a Darlington pair. It serves the same purpose as Q_{A3} in the auxiliary bandgap to improve the current drive in the output V_{ref}. Transistor Q_{R1} is used to bring the current I_1 into the main bandgap. This current is mirrored by the current mirror formed from Q_{M31}, Q_{M32} Q_{M33} into R_4. With this, a PTAT voltage is produced across R_4 and is used to compensate the linear part of V_{BE}. There is a diode-connected transistor M_{R2} in between M_{R1} and Q_{M32}. It is used to minimize the Early effect in the two transistors. The magnitude of the PTAT voltage can be adjusted by proper choice of R_4 to obtain the best possible compensation. In this design, $R_4 = 6.34$ kΩ is chosen.

In addition to the bandgap circuit, there is a start-up circuitry in this design. It is required to pull the bandgap to the desired operating point. Since the auxiliary bandgap is self-biased through the current mirror, it has two stable operating points. One is the desired point that the auxiliary bandgap provides the appropriate output voltage V_{auxref}. The other is the zero current state in which the auxiliary bandgap stays in idle. With the start-up circuit, the auxiliary bandgap is guaranteed to approach the desired operating point after power-up. When the power is first started, a current is developed across the resistor R_{Start}. With the current mirror action, the same current is flowing through Q_{S2}. This current is flowing through M_{A1} and R_1 and R_2 in the auxiliary bandgap. This starts the auxiliary bandgap which in turn find its way to the desired operating point. When the auxiliary bandgap reaches its operating point, transistor Q_{S2} in the start-up circuit will be turned off. The start-up circuit will have no effect to the auxiliary bandgap. Since the value of R_{Start} determines the current level in the auxiliary bandgap during the start-up phase, it has been chosen

to be 40 kΩ to provide a current level that is not too high in the auxiliary bandgap.

The 5 V supply bandgap design is now finished and all design parameters are summarized in Table 4.1. All bipolar transistors in the design are belonged to the NN5111X type provided by the BATMOS library.

Component/Parameter	Value
M	4
R_1	724 Ω
R_2	4 kΩ
R_3	11.3 kΩ
R_4	6.34 kΩ
R_{Start}	40 kΩ
M_{A1}, M_{A2}	40 / 2
M_{A3}, M_{A5}	4 / 2
M_{A4}, M_{A6}	8 / 2
M_{M11}, M_{M12}	40 / 2
M_{P1}	40 / 2
M_{S1}, M_{S3}	40 / 2
M_{S2}	40 / 2
M_{S4}	130 / 2
M_{R1}, M_{R2}	40 / 2
M_{R3}, M_{R4}	8 / 2

Table 4.1: The Design Parameters of the 5 V Supply Bandgap reference

4.3.2 3 V Supply Bandgap Reference Circuit

The proposed compensation technique is also implemented in a design targeted to work with a 3 V supply voltage. This circuit is similar to the 5 V supply bandgap reference in many ways. However, there are some major differences between the two designs. Firstly, the lower supply voltage forces the 3 V design to have fewer components between the supply rails. This ensures all components are operating in linear region. It requires that some of the components in the auxiliary bandgap and the main bandgap circuits from the 5 V design be eliminated so as to fit the 3 V supply

voltage. Secondly, the translinear current polynomial circuit has to be redesigned so that it can operate in a lower supply voltage. Thirdly, the bipolar current mirror and current repeater used in the 5 V design have to be replaced by their NMOS equivalents. The reason is that the bipolar transistors will be saturated in 3 V supply due to the lack of voltage margin. In return, the use of NMOS transistors poses a new constraint in the design. It is required to keep the current levels in the NMOS transistors low in order to minimize the effect of leakage current in the transistors. This requires some changes in the sizings of PMOS transistors. All the above new issues are addressed and all necessary modifications to the 5 V design are described here.

The schematic for the 3 V design is shown in Fig. 4.5. In this figure, it is apparent that many changes are made in this circuit. In the auxiliary bandgap, a number of transistors are removed. The transistors that are used for base current compensation are no longer possible in the 3 V design since they will put the rest of the transistors in the auxiliary bandgap in or close to saturation. Thus, they are discarded. For the same reason, the Darlington pair transistors that are used to boost the output drive of V_{aurref} is now replaced by a single transistor. Since there is no load attached to the output V_{aurref}, a single transistor does not degrade the performance of the auxiliary bandgap. In the main bandgap circuit, a similar measure is necessary. The transistors used for base current compensation in the 5 V design are also removed in the 3 V design. The Darlington output stage is also replaced by a single transistor. Therefore, the current driving capability of the 3 V design will be much worse than its 5 V counterpart. However, this limit is not a serious degradation as voltage reference is always buffered. The removal of base current compensation circuitries in both auxiliary and main bandgap will degrade the

accuracy of the reference voltage as discussed in section 2.6. The second order effect of base current error will be much more prominent. However, the 3 V power supply does not permit their presence. Thus, a worse performance of the 3 V design will be expected. On the other hand, the lower supply voltage also excludes the need for the diode-connected PMOS transistor M_{R2} in the main bandgap circuit. This transistor is used to reduce the Early effect. Since the supply voltage is reduced. Early effect becomes less severe and the transistor is no longer needed.

The second change in the 3 V design is the translinear current polynomial circuit. The main difficulty to transfer the circuit in the 5 V design into 3 V power supply environment comes from the TL loop that generates the current I_4. In this loop. there are four diode-connected transistors (Q_{P1}, Q_{P2}, Q_{P3} and Q_{P4}) in series between the two supply rails. This requires a voltage level of around 2.8 V at room temperature to sustain the four transistors in forward-active mode. With a supply of 5 V. there is no problem. However. there is only 0.2 V left in a supply of 3 V. This is the only voltage left for M_{P1} which brings in I_1. It is definitely not enough for M_{P1} and will force M_{P1} to operate in the ohmic region. Because of this, the current mirror property will be disturbed and the current I_1 will not be mirrored properly into the translinear circuit. As a consequence, a redesign of the translinear circuit is inevitable.

In the new design, the currents I_3 and I_4 are not generated directly from I_1 and I_o. Instead, a $(\text{PTAT})^2$ current I_2 is first generated from I_1 and I_o. After this, I_2 together with I_1 and I_o are then used to generate I_3 and I_4. The current I_2 is generated by the TL loop containing Q_{P1}, Q_{P2}, Q_{P3} and Q_{P4}. By applying the TL principle. the following equation is obtained

$$I_{C1} \cdot I_{C2} = I_{C3} \cdot I_{C4} \tag{4.29}$$

By substituting the corresponding currents, the equation becomes

$$I_1^2 = I_o \cdot I_2 \tag{4.30}$$

Thus, the current I_2 is derived and equal to

$$I_2 = \frac{I_1^2}{I_o} \tag{4.31}$$

This current is used by another two TL loops to generate I_3 and I_4. The current I_3 is generated by the TL loop containing Q_{P5}, Q_{P6}, Q_{P9} and Q_{P10}. By applying the TL principle to the loop currents and after substituting the corresponding currents. the current I_3 is equal to

$$\begin{aligned} I_3 &= \frac{I_2^2}{I_1} \\ &= \frac{I_1^3}{I_o} \end{aligned} \tag{4.32}$$

Similarly, the current I_4 is equal to

$$I_4 = \frac{I_1^4}{I_o^3} \tag{4.33}$$

Thus, both I_3 and I_4 can be generated by this new circuit.

The third required changes in the 3 V design are the bipolar current mirror and current repeater. With a reduced power supply voltage, the bipolar transistors in both the current mirror (Q_{M31}, Q_{M32} and Q_{M33}) and current repeater (Q_{M21}, Q_{M22} and Q_{M23}) can easily be saturated. Therefore, they are replaced by their NMOS equivalents as NMOS transistor does not have this problem. In the case of the current mirror in the main bandgap circuit, it is substituted by a basic two-transistor current mirror. It is chosen to lower the risk of any transistor operating in ohmic region. The three-transistor current mirror has a higher risk due to its requirement of higher headroom voltage which is not available in the 3 V supply voltage. In the case of

current repeater, an exact NMOS equivalent is used. The sizing of all the transistors in both the current mirror and repeater follows those of the PMOS transistors. All have a W:L ratio of 40:2.

Because of the use of NMOS transistors in the 3 V design, the problem of leakage current becomes more severe. At high temperatures, the leakage current in the NMOS transistor will kill the property of current mirror (or repeater). In order to prevent this, a lower operating current level is chosen in the translinear circuit. Therefore, the current I_o and I_1 are scaled down by the mirrors (M_{M11}, M_{M12}) and (M_{MA1},M_{MP1}), respectively. The scale down ratio for I_o is 200:40 while that for I_1 is 40:20. As a result of the scaling down, the currents I_3 and I_4 are scaled up before they are added together by the current summer. It is to compensate the loss in scaling down. The scale up ratio for I_3 is 40:130 while that for I_4 is 40:80. Therefore, the bias current I_C for Q_{R1} is equal to

$$I_C = AT^3 + BT^4 \tag{4.34}$$

where

$$A = \frac{325}{64}\left(\frac{k}{qR_1}\ln M\right)^3\left(\frac{R_3}{V_{aurref}}\right)^2 \tag{4.35}$$

and

$$B = \frac{1000}{64}\left(\frac{k}{qR_1}\ln M\right)^4\left(\frac{R_3}{V_{aurref}}\right)^3 \tag{4.36}$$

This current is slightly different from the one in 5 V design in terms of current ratio of I_3 to I_4 although the two proportions are close to each other. Because of this, the PTAT voltage used for compensation must be different. Here, $R_4 = 5.51$ kΩ is chosen to obtain a different PTAT voltage. The other parameters remain the same as those in 5 V design. All design parameters for the 3 V design are summarized in Table 4.2.

Component/Parameter	Value
M	4
R_1	724 Ω
R_2	4 kΩ
R_3	11.3 kΩ
R_4	5.51 kΩ
R_{Start}	40 kΩ
M_{A1}, M_{A2}	40 / 2
M_{M11}	200 / 2
M_{M12}	40 / 2
M_{P1}	20 / 2
M_{P2}, M_{P3}	40 / 2
M_{S1}, M_{S3}	40 / 2
M_{S2}	80 / 2
M_{S4}	130 / 2
M_{R1}	40 / 2
M_{M21}, M_{M22}, M_{M23}	40 / 2
M_{M31}, M_{M32}, M_{M33}	40 / 2

Table 4.2: The Design Parameters of the 3 V Supply Bandgap reference

4.4 Conclusion

The proposed curvature-compensation technique for bandgap voltage reference is described in this chapter. The theoretical derivation of the new compensation technique is first introduced and the achievable improvement by this new technique is derived. The achievable error is shown to be 60 μV for the temperature range of -55 ^oC to 180 ^oC. The circuit topology for implementing the new compensation technique is given. Two circuits are designed based on this topology, one for a 5 V and the other for a 3 V power supply voltage. The performance of these two circuits are simulated and their results are given in chapter five.

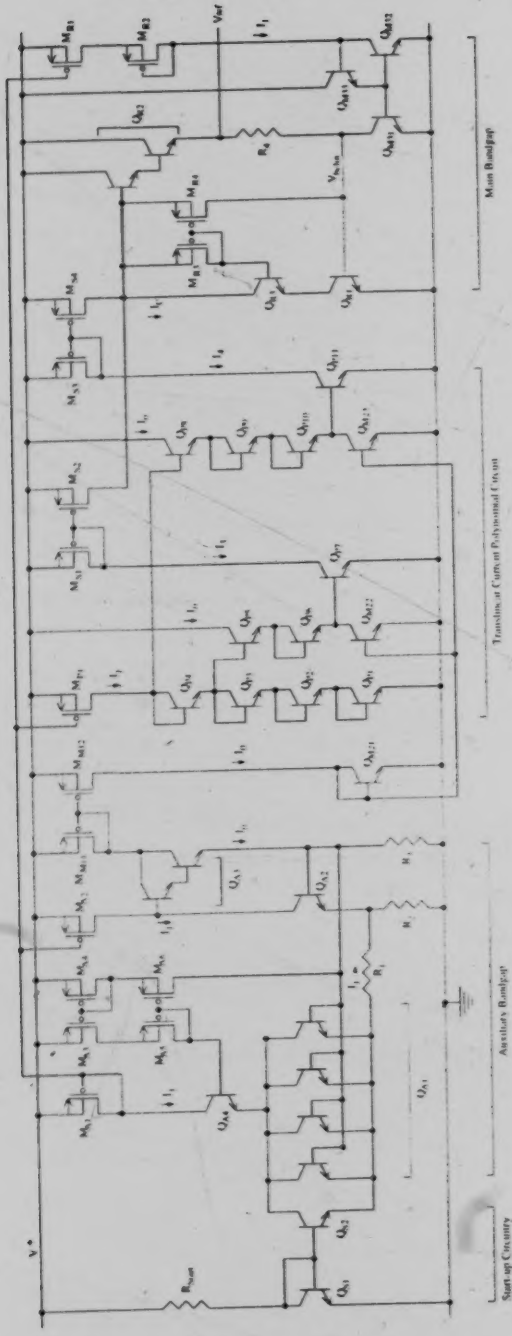

Figure 4.4: Bandgap Reference Circuit for 5 V Supply Voltage

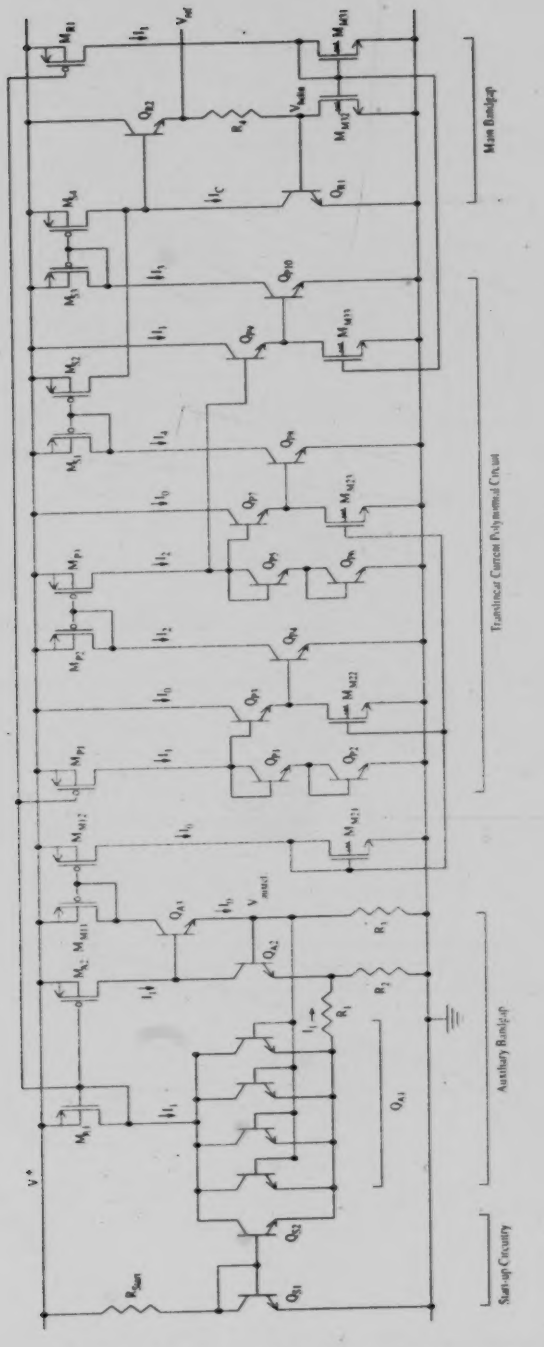

Figure 4.5: Bandgap Reference Circuit for 3 V Supply Voltage

Chapter 5

Simulation Tools and Results

5.1 Introduction

In the previous chapter, two implementations of the proposed curvature-compensation technique for bandgap voltage reference are discussed. In this chapter, the simulation results for the two designs are provided. The simulation results are compared with those of the theoretical prediction. In addition, the linearized base-emitter voltage as well as the voltage in the auxiliary bandgap are studied. A number of bandgap reference parameters for the two designs are also given. Before going into detail of these simulation results, a brief history of circuit simulation is provided. After that, the circuit simulator, Spectre, used for these simulations is described. All the features in Spectre are explored. Then, the simulation results are presented.

5.2 A Brief History in Circuit Simulation

Nowadays, circuit simulation is nearly the first step circuit designers use to verify their designs. In the past, a breadboard scheme was usually employed in which discrete components were used to layout a design. The advantage of this approach is that the exact circuit behavior can be observed. The drawback of this approach is, however, the tediousness of building a large circuit in discrete components. Moreover, this

approach also makes modification to circuit topology more difficult. Because of these reasons, circuit simulation provides a viable alternative for verifying a new circuit design. However, these are not the main trigger to the popularity of circuit simulation. Instead, it is the explosive growth of integrated circuits that makes breadboard prototyping inappropriate. In integrated circuit, prototype is expensive to build and difficult to troubleshoot. Once an integrated circuit has been fabricated, no changes can be made. On the other hand, circuits built from discrete components do not completely represent their behaviors in integrated version as parasitic effects become highly significant. Therefore, it becomes necessary to use simulation such that parasitic effects can be included.

The first well-known circuit simulator is Spice developed by the University of California, Berkeley. The name of Spice is derived from the phrase Simulation Program with Integrated Circuit Emphasis. Spice contains all the models essential for circuit simulation. It is so successful that it becomes the de facto industry standard in circuit simulation. The approach used in Spice has a set of algorithms that find a solution to a circuit in steps. It first formulates the nonlinear ordinary differential equations that describe a circuit. Then, a multi-step integration method like the trapezoidal rule is utilized to convert the differential equations into a system of nonlinear difference equations. After that, Newton-Raphson algorithm is used to solve the nonlinear difference equations. Thus, a sequence of linear equations is generated and is solved by the sparse Gaussian elimination. This approach is commonly called the direct methods which have proven to be the most reliable and general method available.

In the late 70's and early 80's, nearly all integrated circuit manufacturers had their own in-house maintained circuit simulators. All of these simulators are developed based on Spice. The manufacturers have to develop their own simulator because

there was no reliable commercial circuit simulator at that time. This phenomenon started to change in late 80's and early 90's with the emergence of more sophisticated and reliable commercial simulators. These commercial products outperformed the in-house ones in terms of capabilities and efficiency. The companies were economically better off with the use of these commercial simulators than having a team of engineers to manage their own tools. Today, there are many commercial simulators available in the market. For instance, there are AccuSim from Mentor Graphics, PSpice from MicroSim, ViewSpice from ViewLogic Systems. Hspice from Meta-software, and Spectre from Cadence Design. In this thesis, Spectre are used to simulate the two designs. All of its feature are described in next section.

5.3 The Features in Spectre

In this thesis, all circuit designs are simulated using Spectre with device models provided by Northern Telecom based on their 0.8 μm BATMOS process through the aid of Canadian Microelectronic Corporation (CMC). Spectre supports various types of components found essential in circuit simulations. They include: resistors, capacitors, inductors, transformers, delay lines, lossless and lossy transmission lines, microstrip lines, independent voltage and current sources, switches and relays, voltage and current sources that are voltage-controlled or current-controlled, voltage and current sources with polynomial voltage-controlled or current-controlled, voltage and current sources (specified with s-domain or z-domain transfer functions) with linear voltage control or current control, current probes, ports, N-ports described by S-parameter data files, diodes, BJTs, MOSFETs, GaAs MESFETs, JFET and magnetic cores and windings. Together with these components, Spectre can perform the following analyses: DC operating point, transient and small-signal. Within small-signal analysis,

there are AC analysis, small-signal transfer function analysis, S-parameter analysis, time-domain reflectometer analysis and noise analysis. While performing these analyses, Spectre can sweep frequency, temperature or any one of device component parameters.

In DC operating point analysis, Spectre determines the equilibrium point of a circuit by shorting inductors, opening capacitors, and setting all time-varying sources to their quiescent values. For nonlinear devices, Spectre will create linearized models before performing DC analysis. With Spectre, DC transfer curves can be created by specifying a sweep range and a sweep parameter like temperature or a component parameter. DC operating point analysis is also performed before a small-signal analysis to determine the operating point of a circuit.

In transient analysis, Spectre computes the response of a circuit to large input signals specified by the user. The initial state of the circuit can be specified by the user. If there is no initial state specified, the DC operating point will be used. A transient analysis can also be performed while sweeping a parameter like temperature. In this case, the transient analysis will be performed at each step of the sweep parameter.

Spectre can perform a number of small-signal analyses. For each kind of analysis, it will first compute the DC operating point and linearize all nonlinear devices in a circuit. After the circuit is linearized, it will be analyzed over the range of values specified in the independent variable. The independent variable can be the frequency, temperature, or parameter of any component or model. If changing the independent variable affects the DC operating point, Spectre will perform the DC analysis at each value of the variable.

In AC small-signal analysis, Spectre computes the response of a circuit to a small sinusoidal signal. The magnitude of the signal does not affect the circuit response

because all nonlinear devices in the circuit are linearized. Even with this, the signal is normally chosen to have a magnitude of unity and phase of zero. By this way, the small-signal transfer function is directly computed. In transfer function analysis, the transfer function from any node in a circuit to one selected output can be computed. In S-parameter analysis, the S-parameters between one or more ports in a circuit can be computed. These S-paramters are used to describe a linear N-port network and can be written to a data file such that they can be used in another analysis. In time-domain reflector analysis, Spectre computes the reflection coefficients versus time, looking into a circuit from its input ports. This analysis is the time-domain equivalent of S-parameter analysis. In noise analysis, all the noise contributions of individual components in a circuit are computed as total output noise. The total noise contributions can also be computed as equivalent input noise at the selected input of a circuit.

5.4 Simulation Results of the 5 V Design

A number of different kinds of simulations are done for the 5 V design of the proposed bandgap reference. The first simulation done for the circuit is the steady state response of the design to temperature variations. In this simulation, a 5 V dc power source is used to power up the circuit. A transient analysis is performed for the circuit while sweeping temperature. The temperature range for sweeping is -55 ^oC to 180 ^oC. A step interval of 5 ^oC is used and the transient analysis is performed at each temperature step for a time of 2 ms. The observed time at each temperature step is 1 ms. The output voltage V_{ref} of the bandgap reference for each temperature step at time 1 ms is plotted in Fig. 5.1. From the figure, the maximum variation of V_{ref} is seen to be 1.4 mV. Therefore, using eqn. (1.2), the temperature coefficient

of the 5 V supply bandgap reference is 5 $ppm/°C$. This variation is much greater than that predicted by the theoretical calculation in section 4.2. It is probably due to many second order effects that are not considered in the theoretical calculation. The linearized base-emitter voltage in the main bandgap is illustrated in Fig. 5.2. As can be seen, this voltage is now better linearized. The voltage V_{auxref} in the auxiliary bandgap is shown in Fig. 5.3. This voltage has a maximum variation of 6 mV and is close to the theoretical prediction. In addition, the output driving capability of the circuit is analyzed. The output of the bandgap is used to drive an resistive load R_{load} which is swept from the values of 1 $k\Omega$ to 20 $k\Omega$. The analysis is done at the default temperature of Spectre, namely 27 $°C$. The result is demonstrated in Fig. 5.4. As can be seen, there is not significant change in V_{ref} for values of R_{load} from 10 $k\Omega$ to 20 $k\Omega$. It only varies by less than 0.2 mV. However, its value deteriorates quickly when the resistor value goes down from 10 $k\Omega$ to 1 $k\Omega$. If the value of R_{load} is going down even further, the value of V_{ref} will be even lower. Therefore, this bandgap reference cannot be loaded heavily. For R_{load} values between 10 $k\Omega$ and 20 $k\Omega$, load regulation can be found. The output voltage at R_{load} of 10 $k\Omega$ is 1.11848 mV while that at 20 $k\Omega$ is 1.11861 mV. Therefore, the load regulation of the 5 V supply reference is -2.3 $\mu V/\mu A$. Another interesting response of the bandgap reference is its transient response to start-up. In order to illustrate this, a 5 V step input with a rise time of 1 μs is used to power up the bandgap circuit. The response of the circuit is shown in Fig. 5.5. It can be seen that the response of the circuit follows close to the supply input and there is a glitch in its rising. Moreover, the ripple rejection ratio (RRR) of the bandgap circuit is also of interest. A sinusoidal voltage with peak-to-peak voltage of 0.2 V and dc voltage of 5 V is used to find the response of the circuit. The peak-to-peak variation of the circuit can be observed and the RRR can be found by

the following equation

$$RRR = -20Log\left\{\frac{V_{refpp}}{V_{supplypp}}\right\} dB \qquad (5.1)$$

This equation is derived from eqn. (1.3) in chapter 1. The response of the circuit is shown in Fig. 5.6. From the figure, it is shown that the peak-to-peak variation of V_{ref} is 0.01 V. Therefore, the RRR is 26 dB. Although it is not a very good figure, it is not very important as bandgap reference is seldom used as a stand-alone circuit. However, a more sophisticated design is necessary when the RRR becomes critical. Fig. 5.6 only shows that the response of the circuit at 27 ^oC. The output voltage V_{ref} in response to a supply voltage of 5.1, 5.0 and 4.9 V for the temperature range of interest is shown in Fig. 5.7. The variations of V_{ref} at different supply voltages can be seen from this figure. The line regulation of the reference can be derived from this figure for the default temperature of 27 ^oC. At the supply voltage of 5.1 V, the output voltage is 1.1239 V at 27 ^oC. On the other hand, the output voltage at 27 ^oC is 1.1136 V for the supply voltage of 4.9 V. Therefore, the line regulation is 51.5 mV/V. All of these parameters are summarized in Table 5.1.

Parameter	Value
Temperature Coefficient	5 ppm/^oC (1.4 mV)
Load Regulation	-2.3 μV/μA
Ripple Rejection Ration	26 dB
Line regulation	51.5 mV/V

Table 5.1: The Performance Parameters of the 5 V Supply Bandgap reference

5.5 Simulation Results of the 3 V Design

The same set of simulations is also performed for the 3 V supply bandgap reference. The response of the output voltage V_{ref} to the temperature variations of -55 ^oC to 180

$°C$ is shown in Fig. 5.8. The maximum variation is about 2.3 mV which corresponds to a temperature coefficient of 9 $ppm/°C$. It is not as good as the 5 V circuit, however, this result is expected since a lot of the second order effect minimization features are removed in the 3 V design. The linearized V_{BE} of the main bandgap is shown in Fig. 5.9. It has a less smooth appearance than that of the 5 V circuit. The voltage V_{auxref} in the auxiliary bandgap is demonstrated in Fig. 5.10. The maximum variation is equal to 15 mV which is somewhat away from the theoretical prediction. It is mainly due to the error in base current as the base current compensation circuit is removed. The current driving capability of the 3 V circuit can be seen in Fig. 5.11. Similar to the 5 V circuit, it does not have a significant variation for values of R_{load} from 10 $k\Omega$ to 20 $k\Omega$. The value of V_{ref} goes down with decreasing R_{load} starting at 10 $k\Omega$. For R_{load} between 10 $k\Omega$ and 20 $k\Omega$, the load regulation is found to be -2.74 $\mu V/\mu A$. The transient response of the circuit to a 3 V step of 1 μs rise time is illustrated in Fig. 5.12. Its response follows close to the input supply voltage. The response of the reference to a sinusoidal of 0.2 V peak-to-peak and 3 V dc at 27 $°C$ is depicted in Fig. 5.13. The peak-to-peak voltage of V_{ref} is 12 mV. By using eqn. (5.1), the RRR is equal to 24 dB. Finally, the response of the output voltage V_{ref} to a supply voltage of 5.1, 5.0 and 4.9 V for the whole temperature range of interest is demonstrated in Fig. 5.14. The line regulation at 27 $°C$ can be found to be 57.5 mV/V. All of these parameters are summarized in Table 5.2.

5.6 Conclusion

In this chapter, a brief review in the history of circuit simulation is provided. The circuit simulator used in this thesis, namely, Spectre is described and all of its features are explained. The simulation results of both the 5 V and 3 V supply bandgap

Parameter	Value
Temperature Coefficient	9 ppm/^oC (2.3 mV)
Load Regulation	-2.74 μV/μA
Ripple Rejection Ration	24 dB
Line regulation	57.5 mV/V

Table 5.2: The Performance Parameters of the 3 V Supply Bandgap reference

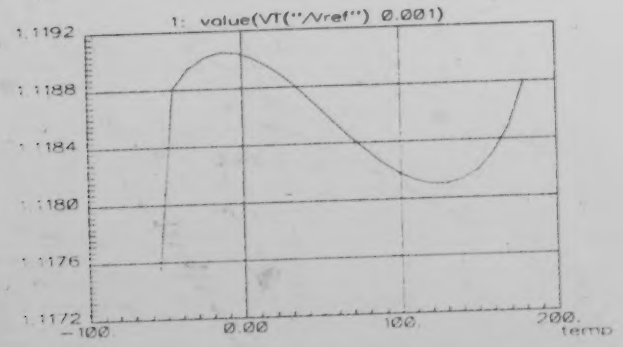

Figure 5.1: The Reference Voltage of the 5 V Design versus Temperature

references are provided and the performances of the two designs are discussed. The temperature coefficients of both references are found to be quite low for the temperature range of -55 ^oC to 180 ^oC. It is considered a significant improvement over previous designs. These prove that the proposed compensation technique can provide a reference with better temperature stability.

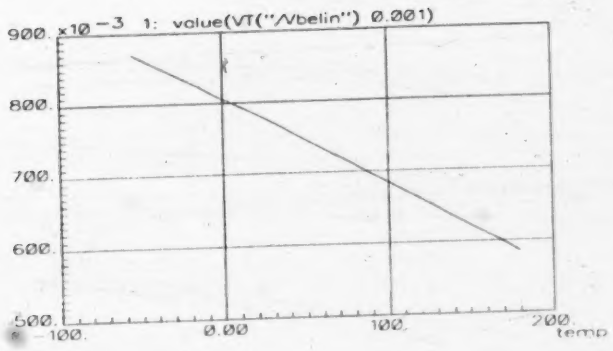

Figure 5.2: The Base-emitter Voltage of the 5 V Design versus Temperature

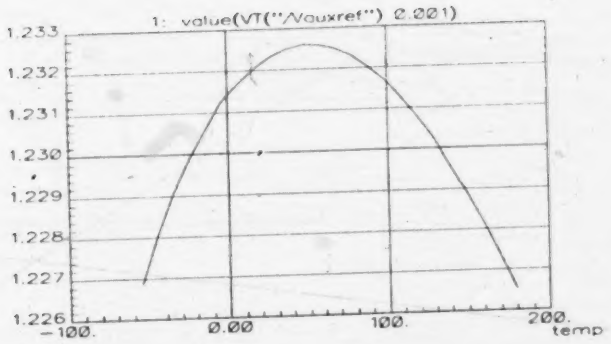

Figure 5.3: The Auxiliary Bandgap Voltage of the 5 V Design versus Temperature

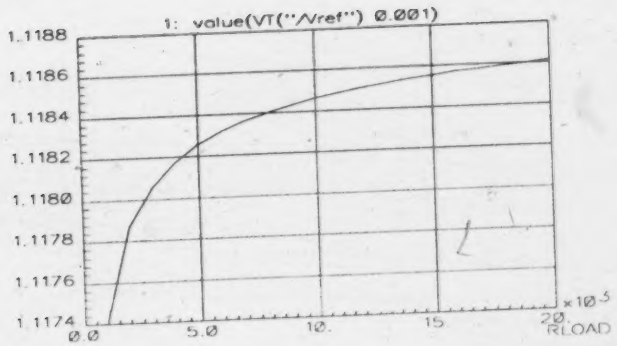

Figure 5.4: The Reference Voltage of the 5 V Design for Various Loads at 27 ^{o}C

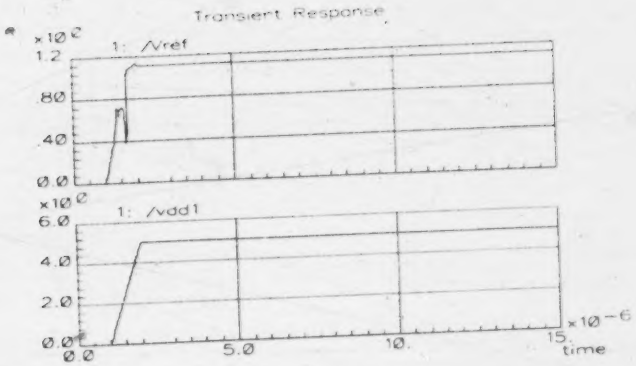

Figure 5.5: The Transient Response of the 5 V Design at 27 ^{o}C (Input Rise Time of 1 μs)

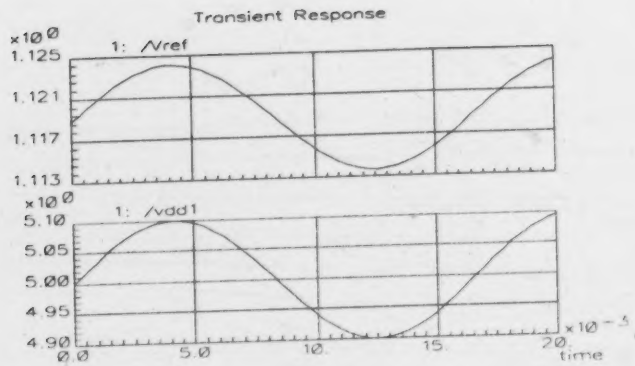

Figure 5.6: The Response of the 5 V Design to a Supply Voltage with 0.1 V of Variations (at 27 ^{o}C)

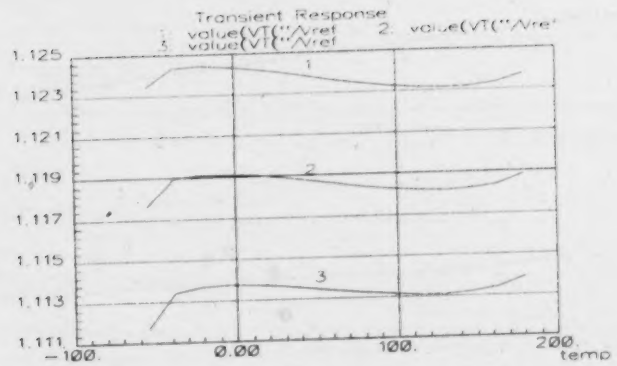

Figure 5.7: The Response of the 5 V Design to a Supply Voltages of 5.1, 5.0 and 4.9 V

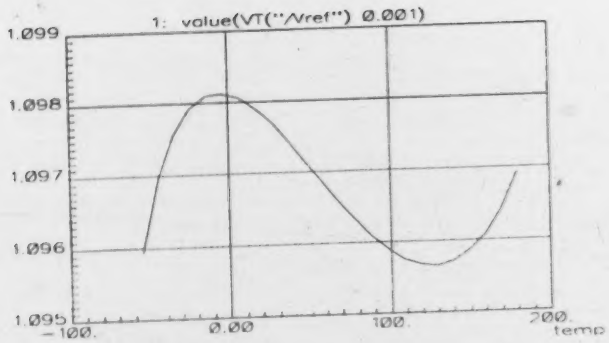

Figure 5.8: The Reference Voltage of the 3 V Design versus Temperature

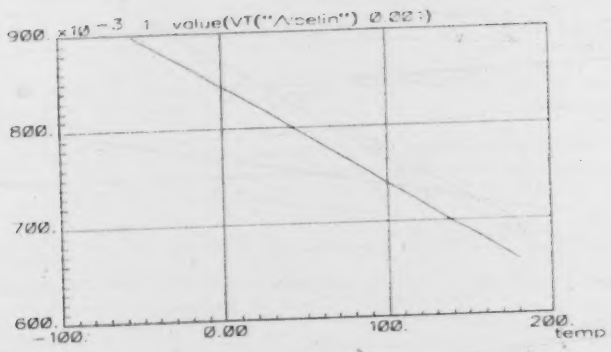

Figure 5.9: The Base-emitter Voltage of the 3 V Design versus Temperature

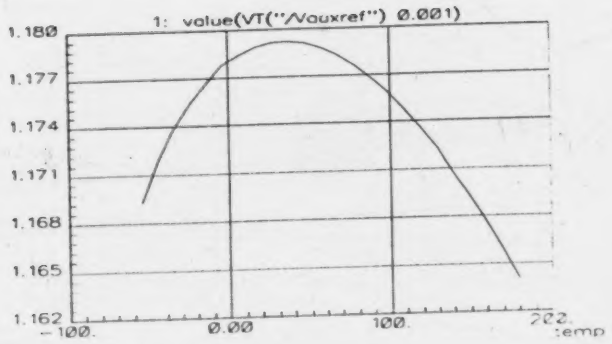

Figure 5.10: The Auxiliary Bandgap Voltage of the 3 V Design versus Temperature

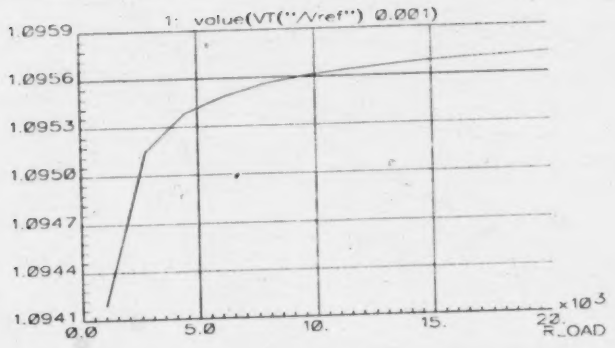

Figure 5.11: The Reference Voltage of the 3 V Design for Various Loads at 27 $^{\circ}C$

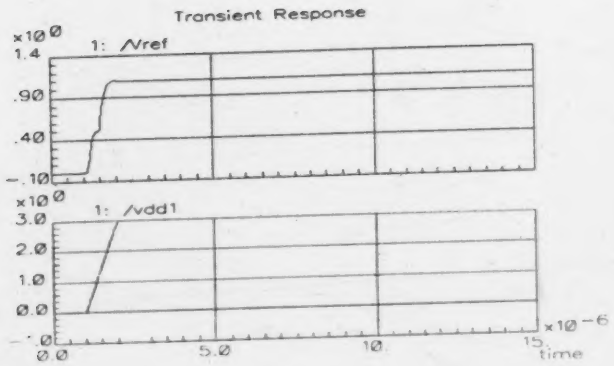

Figure 5.12: The Transient Response of the 3 V Design at 27 ^{o}C (Input Rise Time of 1 μs)

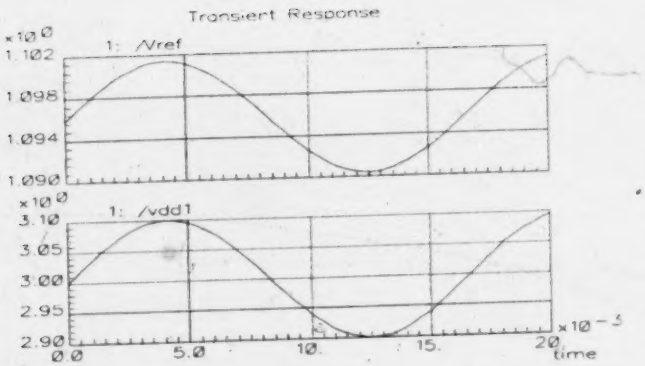

Figure 5.13: The Response of the 3 V Design to a Supply Voltage with 0.1 V of Variations (at 27 ^{o}C)

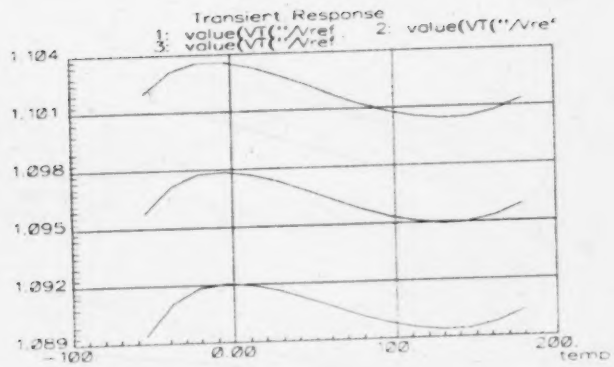

Figure 5.14: The Response of the 3 V Design to a Supply Voltages of 3.1, 3.0 and 2.9 V

Chapter 6

Layout Design

6.1 Introduction

Before any circuit design can be fabricated in any given technology, it must be translated from transistor level to layout level representation. For digital circuit, there are many tools that provide automatic translation. However, the same kind of tools is not available for analog circuits. In fact, it is because an analog circuit usually requires more sophisticated component placement and signal routing that makes automatic translation impossible. Therefore, circuit designers usually have to manually perform the translation for analog circuits. In spite of this inconvenience, there is still a tool that aids designers to translate a circuit from a schematic to a layout. This tool is known as *layout editor* which allows a designer to put different polygons on various mask layers like diffusion or metal layers to obtain a functional physical representation of a given design. In addition to aiding layout, a layout editor is usually integrated with a design rule checker. This rule checker is used to ensure that a given layout does not violate any design rule as defined by the target technology used for implementing the design. This set of design rules is normally provided in a technology file such that the design rule checker can be utilized to verify that a given layout complies to all the design rule requirements.

Having a layout without any design rule violation does not guarantee it produces a circuit with the same functionality as the design in transistor level. In order to check whether a given layout is functionally the same as its transistor level representation, a designer can use a tool called *circuit extractor* to derive a circuit schematic from a physical layout. The extractor reconstructs a transistor level representation from a physical layout by scanning various layers and their interactions. The reconstructed circuit includes the sizes of the devices and their interconnections. The circuit designer can then perform the same type of simulations used in the transistor level circuit to verify that the layout is correctly representing the desired circuit.

In addition to extracting a transistor level representation from a layout, a circuit extractor can also include information on parasitics in the extracted circuit. This information can be diffusion and wiring capacitances as well as resistances. The inclusion of this information helps the designer to forecast the effect of parasitics in actual circuit. The designer can then decide whether certain parasitic effect is severe and make appropriate modifications to the layout. By doing this, the resulting layout has a better chance of providing the desire functionality as designed in the transistor level schematic.

The previous chapter shows the simulation results of the two proposed bandgap references in transistor level and proves that the new approach can provide a better performance compared to previous bandgap reference designs. In this chapter, the translations of the two designs from transistor level to layout level are provided. These two layouts are extracted for simulation. The temperature dependencies of the extracted circuits are demonstrated.

6.2 Layout Considerations

In translating a circuit from transistor level to layout level, there are some general rules to follow for minimizing parasitic effect. The commonly occurring parasitics in integrated circuits can be classified into three main types, namely, capacitive, resistive, and inductive parasitics. For capacitive parasitics, the main contributions are caused by interconnect wires. These interconnect wires form capacitance together with the insulating layer and the substrate. These parasitics will delay the transmission of a signal from one point of a circuit to another. This is because it takes time for signal to charge and discharge the capacitance. In order to minimize these parasitics. it is necessary to keep the dimension of the routing interconnect wires as small as possible. Therefore, it is desirable to have the interconnect as small as the technology allows. The other source of parasitic capacitance is the coupling between two signaling wires. This coupling results in interference between the two wires and is often referred to as cross talk. This kind of parasitics can be prevented by avoiding the routing of parallel signal lines. It can also be minimized by increasing the separation of the two signal wires. The aforementioned capacitive parasitics are the two major ones usually considered by a designer. The same types of preventive measures are followed in this thesis.

The second type of parasitics a designer will consider after capacitive parasitics is resistive parasitics. It arises mainly from two sources. The first one is from the interconnect wires in an integrated circuit. The effect is that this resistance causes an ohmic voltage drop that degrades the integrity of a signal. The severity of this effect depends on the nature of the material used for the routing. For polysilicon, it has a larger sheet resistance compared to metal. Thus, it is generally not a good

choice to use polysilicon for long routing layer. However, the use of polysilicon as routing material is acceptable if the length of the routing is limited to localized signals. Therefore, a circuit designer usually uses metal layer if a long routing layer is necessary. The diffusion layer is always not a good choice for routing because of its large associated parasitic capacitance. The second major source of resistive parasitics comes from the contact or via which provides the transition between two routing layers. They can be minimized by increasing the contact size. However, the size cannot be increased as much as desired due to the current crowding effect where current tends to concentrate around the perimeter of a large contact hole. This puts a practical limit to increase the size of a contact, not to say the feasibility of the size in integrated circuit. In spite of this fact, there are two solutions to get around this problem. The first one is to use multiple smaller contacts instead of one single large one. This will minimize the current crowding problem while decrease the resulting resistance of the transition. The second solution is by trying to limit to one routing layer as the layout strategy. All these precautions are observed in the two layouts designed.

The last type of parasitics is inductive parasitics. This parasitics is more prominent in bonding wires and chip packaging pins. Inductors have the property of resisting current change through them by raising the voltage drop across them. This property will cause a problem when a large signal switching is happening in an I/O pin of a digital circuit where its power is supplied by a power supply pin. When a large signal switching occurs in an I/O pin, a significant amount of current is drawn from the power pin. This will trigger the inductance in this power pin to increase the voltage drop across it. Therefore, an instantaneous drop of supply voltage level for other part of the circuit will happen. This will affect the proper operation of this

part of the circuit. There are a number of approaches a designer can use to address this problem. The most common practice is to use separate power pins for I/O pads and chip core. This can isolate the largest switching currents in the off-chip pins from the core circuit. Another approach is to use multiple power and ground pins. This helps to reduce the number of simultaneous switching actions occurring in one power pin. The third approach is to avoid using power pins at the corners of a package where inductance is substantially higher as the bonding wire is longer. The inductive parasitics is not critical for bandgap reference design since there is not large signal switching action. Therefore, the inductive parasitics effect is not considered in this thesis. However, separate power pads for I/O pins and core circuit are still used.

In addition to the above guidelines, there is one general rule to follow in designing a bandgap reference layout. From the theoretical derivation of the new bandgap reference design in chapter 4, it is shown that the accuracy of this new design depends on the matching and tracking of both the transistors and the resistors. The matching of these devices depends on the fabrication and a circuit designer cannot do much about it. However, a designer can improve the tracking of devices by means of careful layout. In order to ensure a better tracking, it is necessary to avoid thermal gradient for the temperature-sensitive parts of the circuit. The first part is the four resistors (R_1, R_2, R_3, and R_4). These resistors are mainly responsible for setting the voltage level of the bandgap reference. Therefore, ensuring they can operate in a close thermal environment is important. It can be achieved by putting these resistors close to one another in the layout. The second part is the five transistors (Q_{A1} and Q_{A2}) in the auxiliary bandgap. Since these transistors are responsible for generating the temperature-constant and temperature varying currents, I_o and I_1, the tracking of them is very important to the subsequent accuracy of the reference. It is ensured by

locating them close to one another in the layout. The last part is the transistors in the translinear current polynomial circuit. As this part of the reference is used to generate the currents, I_3 and I_4 and the principle of translinear circuit highly dependent on the tracking of transistors, it is obvious that placing all these transistors close to one another in the layout is needed for avoiding the thermal gradient problem.

When a circuit layout is ready for fabrication, a designer must add I/O pads to the layout for signals that go off-chip. These pads contain special circuitry which provides the appropriate driving capability for the off-chip signals. When the pads are included in the layout, this pad circuitry can be extracted so that the designer can predict the performance of the whole circuit including the I/O pads. Provided with the BiCMOS technology library, there are a number of I/O pads for analog circuit. These pads can be divided into two main groups. One has electrostatic discharge (ESD) protection while the other does not. The ESD-protected pad has extra circuitry in place such that it can withstand a large amount of voltage surge without damaging the internal circuit. However, this protection circuitry imposes an upper temperature bound for the whole circuit to operate because it introduces the leakage current problem at high temperature. Within each group of I/O pads, there are a number of different pads serving for different purposes. These include general-purpose I/O pad, core circuitry power pad, core circuitry ground pad, pad power pad, and pad ground pad. The general-purpose I/O pad is used for signal that goes off-chip. The core circuitry power and ground pads serve to bring in power and ground for the internal circuit. The pad power and ground pads do the same job as their core counterpart but they serve the pads only. In this chapter, both the ESD-protected and non-ESD-protected pads are included in the 5 V and 3 V bandgap reference circuit layout. The effect of them can then be demonstrated by simulation.

6.3 Simulation Results

Fig. 6.1 shows the output voltage of the 5 V supply bandgap reference with non-ESD-protected pads. It can be seen that this extracted circuit from the layout shows a similar performance as its transistor level counterpart in chapter 5. The linearized base-emitter voltage in Fig. 6.2 as well as the auxiliary bandgap voltage in Fig. 6.3 also show a similar temperature variation as those of the transistor level in Fig. 5.2 and Fig. 5.3. It follows that two conclusions can be implied from these results. The first is that the layout provides the same circuit as the transistor level schematic. The second one is that the I/O pads do not introduce any differences to the circuit.

The output voltage of the 5 V supply bandgap reference with ESD-protected pads is illustrated in Fig. 6.4. This voltage starts to show a drastic difference from that in Fig. 6.1 when the temperature reaches about 130 ^oC. This is due to the extra protection circuitry in the I/O pads. However, both the linearized base-emitter voltage in Fig. 6.5 and the auxiliary bandgap voltage in Fig. 6.6 do not show any significant deviation.

Fig. 6.7 depicts the output voltage of 3 V supply bandgap reference with non-ESD-protected pads. The performance of this extracted circuit is similar to that of the transistor level shown in Fig. 5.8. The linearized base-emitter voltage is illustrated in Fig. 6.8. This voltage also shows a close similarity to that of the transistor level in Fig. 5.9. The same is true for the auxiliary bandgap voltage of the extracted layout shown in Fig. 6.9. This voltage also shows a similar temperature variation as that in Fig. 5.10.

The output voltage of the 3 V supply bandgap reference with ESD-protected pads is depicted in Fig. 6.10. As that of the 5 V supply bandgap reference, it shows a

significant difference from that in Fig. 6.7 when the temperature exceeds 130 ^{o}C. Again, this is due to the different I/O pads used in the layout. Both the linearized base-emitter voltage and the auxiliary bandgap voltage in Fig. 6.11 and Fog. 6.12, however, do not show the same behavior.

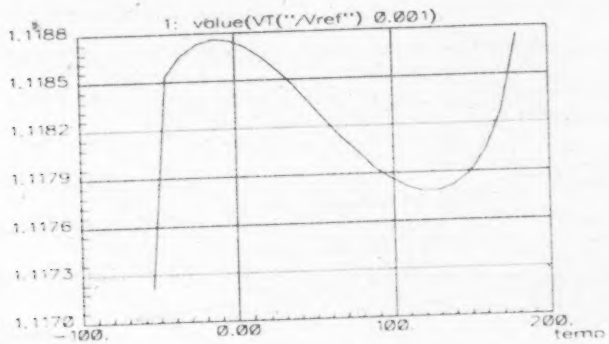

Figure 6.1: The Output Voltage of the 5 V Supply Bandgap Reference with Non-ESD-protected Pads

6.4 Conclusion

In this chapter, the tools available for an analog circuit designer to translate a design from a transistor schematic to a layout representation are discussed. These tools are known as layout editor and circuit extractor. A layout editor can put various polygons together to represent a layout as well as verify a layout does not violate any design rule defined in a given technology. A circuit extractor, on the other hand, can derive a transistor level circuit from a layout such that a designer can verify that the layout is correctly representing the original design through simulation. The extractor can also include additional information like parasitics to the extracted circuit. This

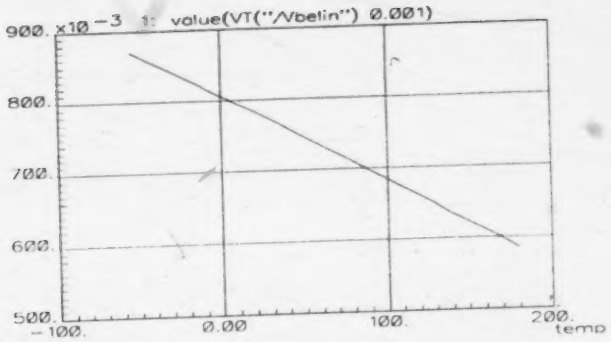

Figure 6.2: The Linearized Base-emitter Voltage of the 5 V Supply Bandgap Reference with Non-ESD-protected Pads

can reveal a lot of useful information for predicting how the final circuit will behave. It is followed by precaution steps necessary to reduce the effect of parasitics in the two designed layouts for both the 5 V and 3 V supply bandgap references in chapter four. Finally, the simulation results of the two layouts are provided. The results demonstrate that when non-ESD-protected I/O pads are included in the layouts they behave similarly as those in transistor level. However, the simulation results of the layouts with the ESD-protected I/O pads show significant difference for the output of the reference when the temperature exceeds 130 ^{o}C.

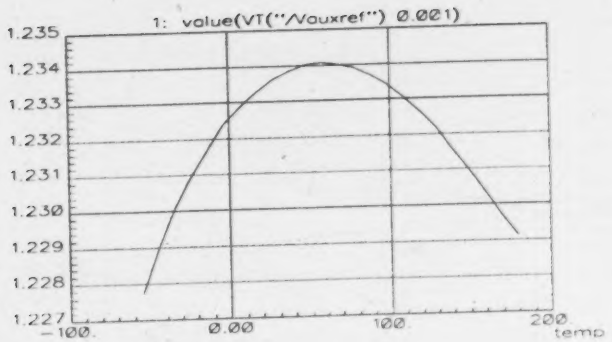

Figure 6.3: The Auxiliary Bandgap Voltage of the 5 V Supply Bandgap Reference with Non-ESD-protected Pads

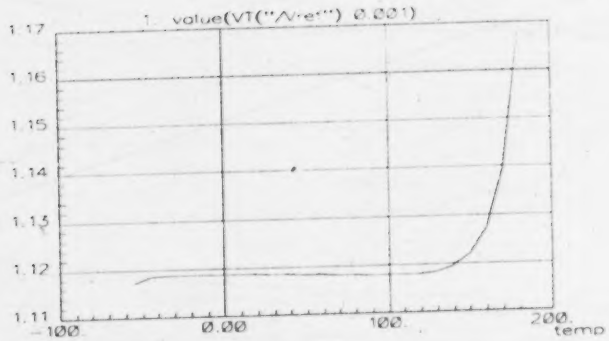

Figure 6.4: The Output Voltage of the 5 V Supply Bandgap Reference with ESD-protected Pads

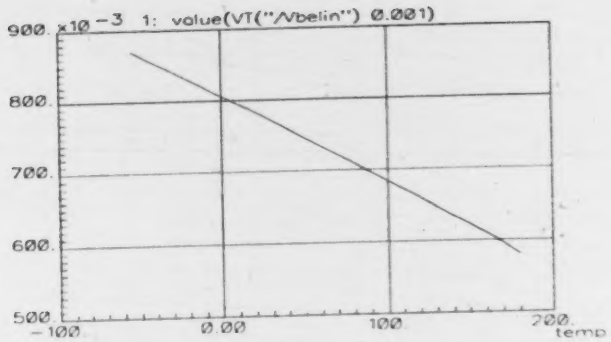

Figure 6.5: The Linearized Base-emitter Voltage of the 5 V Supply Bandgap Reference with ESD-protected Pads

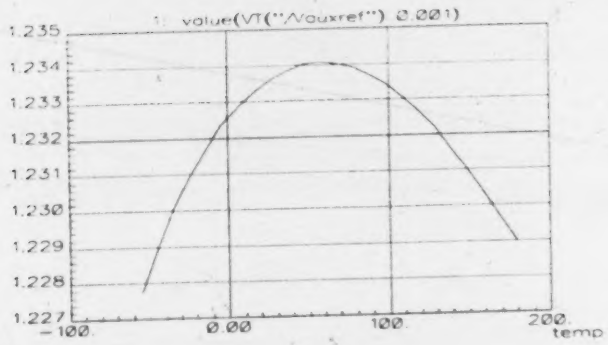

Figure 6.6: The Auxiliary Bandgap Voltage of the 5 V Supply Bandgap Reference with ESD-protected Pads

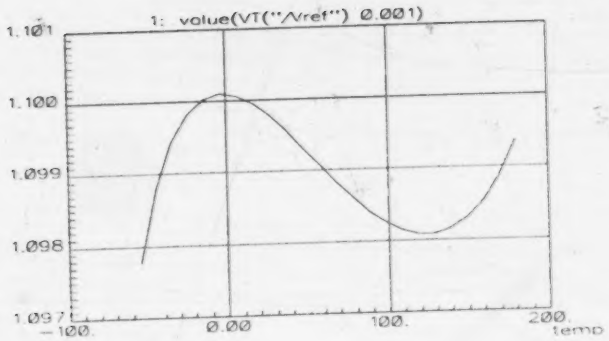

Figure 6.7: The Output Voltage of the 3 V Supply Bandgap Reference with Non-ESD-protected Pads

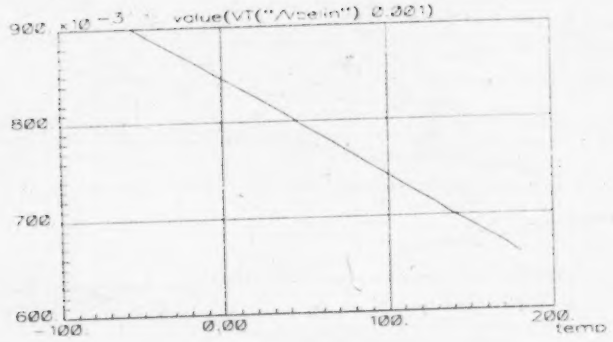

Figure 6.8: The Linearized Base-emitter Voltage of the 3 V Supply Bandgap Reference with Non-ESD-protected Pads

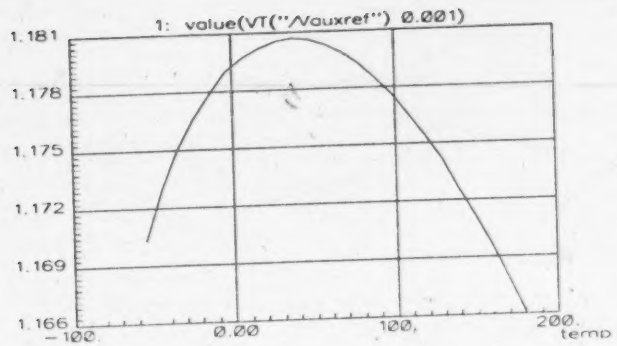

Figure 6.9: The Auxiliary Bandgap Voltage of the 3 V Supply Bandgap Reference with Non-ESD-protected Pads

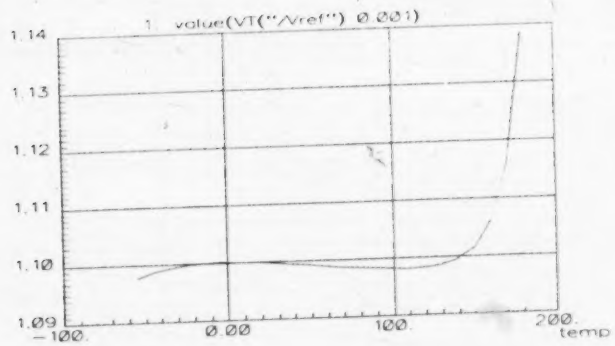

Figure 6.10: The Output Voltage of the 3 V Supply Bandgap Reference with ESD-protected Pads

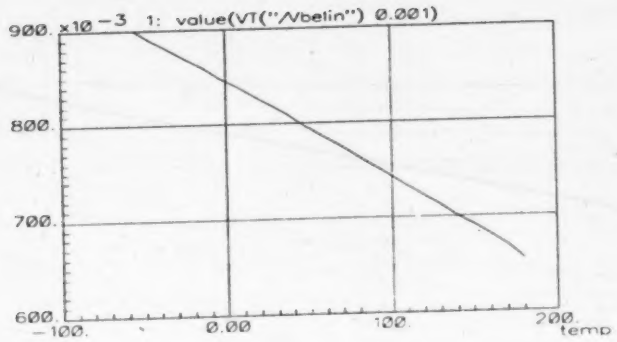

Figure 6.11: The Linearized Base-emitter Voltage of the 3 V Supply Bandgap Reference with ESD-protected Pads

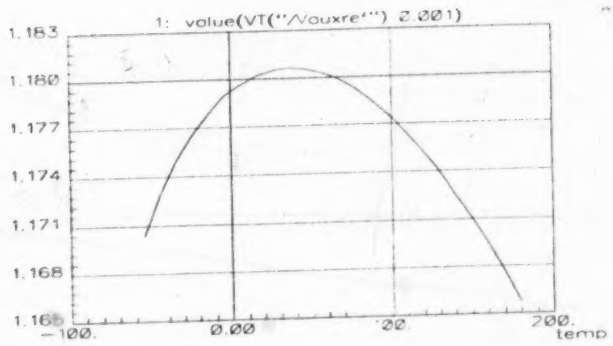

Figure 6.12: The Auxiliary Bandgap Voltage of the 3 V Supply Bandgap Reference with ESD-protected Pads

Chapter 7

Conclusion

In this thesis, various kinds of voltage reference are studied. Bandgap voltage reference is shown to be the best candidate for both low power and low voltage applications. The problem with this kind of reference, however, is the nonlinear temperature-dependent term associated with the base-emitter voltage of a bipolar transistor normally used in the bandgap reference design. There have been a lot of proposed methods in the past for compensating this nonlinear term. The curvature-compensation technique proposed here is to use a proportional sum of two currents, one $(PTAT)^3$ and one $(PTAT)^4$, as bias current to the transistor. This approach provides a better linearized base-emitter voltage. The theoretical prediction shows that this technique provides a bandgap reference with only 60 μV variation for the temperature range of -55 ^{o}C to 180 ^{o}C. In order to implement this technique, the bipolar translinear circuit principle is introduced. Based on this principle, a special circuit is designed for use in the proposed bandgap reference design. This circuit is called translinear current polynomial circuit and is utilized to generate the $(PTAT)^3$ and $(PTAT)^4$ currents. Two versions of this circuit are designed. One is for 5 V supply voltage application and the other for 3 V application. The basic circuit topology for the whole bandgap reference circuit is presented in chapter four and two versions are designed out of it.

These two circuits are designed for two different power supply voltage, one for 5 V and one for 3 V. Both of them show a promising improvement through simulation. The simulation result of the 5 V supply design shows a maximum variation of 1.4 mV (5 ppm/^oC) while that of the 3 V supply design shows 2.3 mV (9 ppm/^oC) over the temperature range of -55 ^oC to 180 ^oC. The layouts for the two designs are also provided and both of them also show a similar performance corresponding to their transistor level counterparts. Although these results are not as low as the theoretical prediction of 60 μV, they are already a significant improvements for such a temperature range. This work is also published in [31].

There are two major areas that can be investigated in the future. The first one is to identify the second order effect that makes the performance of the two circuits presented here away from the theoretical derivation. There are a few places to be focused. Firstly, the constant current source derived from the auxiliary bandgap circuit is not a highly constant current. This effect can be explored by including the exact expression of this current into the theoretical derivation of the proposed technique. The effect of ignoring the non-constant term in this current source can then be shown. Secondly, the accuracy of the $(PTAT)^3$ and $(PTAT)^4$ currents generated by the translinear current polynomial circuit can be investigated. Thirdly, any possible second order effects that might present in the main bandgap circuit can be identified. If any of the above three shows prominent effect, appropriate adjustments can then be made.

The second area to investigate is the use of a proportional sum of $(PTAT)^{3.25}$ and $(PTAT)^{3.75}$ currents as bias current to a bipolar transistor. With these two currents, the achievable error should be smaller. It is not difficult to develop circuit to generate these currents based on the bipolar translinear circuit principle. These two currents

should, in theory, provide a better linearized base-emitter voltage based on the current finding.

When all of the aforementioned improvements are taken into account, it should be possible to develop a bandgap voltage reference with a fraction of ppm/^oC performance.

Bibliography

[1] David P. Laude and James D. Beasom. 5 V temperature regulated voltage reference. *IEEE Journal of Solid-State Circuits*, SC-15(6):pages 1070–1075, December 1980.

[2] Masanori Otsuka, Akira Hyogo, and Keitaro Sekine. Temperature compensation methods for junction-isolated analog integrated circuits and their applications. *Electronics and Communications in Japan, Part 2*, 75(4):pages 107–118, 1992.

[3] Richard J. Reay, Erno H. Klaassen, and Gregory T. A. Kovacs. A micromachined low-power temperature-regulated bandgap voltage reference. *IEEE Journal of Solid-State Circuits*, 30(12):pages 1374–1381, December 1995.

[4] G. C. M. Meijer. Thermal sensors based on transistors. In S. Moddlehoek and J. van der Spiegel, editors, *Sensors and Actuators*, volume 10, pages 103–125. Elsevier, Lausanne, 1987.

[5] Canadian Microelectronics Corporation, Ottawa. *CMC BiCMOS Design Kit V1.0 for Cadence Analog Artist*, 1995.

[6] Paul R. Gray and Robert G. Meyer. *Analysis and Design of Analog Integrated Circuits*. John Wiley & Sons, New York, 3rd edition, 1993.

[7] R. J. Widlar. New developments in IC voltage regulators. *IEEE Journal of Solid-State Circuits*, SC-6(1):pages 2–7, February 1971.

[8] K. E. Kuijk. A precision reference voltage source. *IEEE Journal of Solid-State Circuits*, SC-8(3):pages 222–226, June 1973.

[9] A. P. Brokaw. A simple three-terminal IC bandgap reference. *IEEE Journal of Solid-State Circuits*, SC-9(6):pages 388–393, December 1974.

[10] Yannis P. Tsividis and Richard W. Ulmer. A CMOS voltage reference. *IEEE Journal of Solid-State Circuits*, SC-13(6):pages 774–778, December 1978.

[11] G. Tzanateas, C. A. T. Salama, and Y. P. Tsividis. A CMOS bandgap voltage reference. *IEEE Journal of Solid-State Circuits*, SC-14(3):pages 655–657, June 1979.

[12] Eric A. Vittoz and Olivier Neyroud. A low-voltage CMOS bandgap reference. *IEEE Journal of Solid-State Circuits*, SC-14(3):pages 573–577, June 1979.

[13] Robert A. Blauschild, Patrick A. Tucci, Richard S. Muller, and Robert G. Meyer. A new NMOS temperature-stable voltage reference. *IEEE Journal of Solid-State Circuits*, SC-13(6):pages 767–774, December 1978.

[14] Ho-Jun Song and Choong-Ki Kim. A temperature-stabilized SOI voltage reference based on threshold voltage difference between enhancement and depletion NMOSFET's. *IEEE Journal of Solid-State Circuits*, 28(6):pages 671–677, June 1993.

[15] Alan B. Grebene. *Bipolar and MOS Analog Integrated Circuit Design.* John Wiley & Sons, New York, 1984.

[16] Eric A. Vittoz. MOS transistors operated in the lateral bipolar mode and their application in CMOS technology. *IEEE Journal of Solid-State Circuits*, SC-18(3):pages 273–279, June 1983.

[17] Bang-Sup Song and Paul R. Gray. A precision curvature-compensated CMOS bandgap reference. *IEEE Journal of Solid-State Circuits*, SC-18(6):pages 634–643, December 1983.

[18] John Michejda and Suk K. Kim. A precision CMOS bandgap reference. *IEEE Journal of Solid-State Circuits*, SC-19(6):pages 1014–1021, December 1984.

[19] A. van Staveren, C. J. M. Verhoeven, and A. H. M. van Roermund. The influence of the reverse early effect on the performance of bandgap references. *IEEE Transactions on Circuits and Systems - Part I*, 43(5):pages 418–421, May 1996.

[20] O. Salminen and K. Halonen. The higher order temperature compensation of bandgap references. *IEEE International Symposium on Circuits and Systems*, 3:pages 1388–1391, May 1992.

[21] Jonathan M. Audy. 3rd order curvature corrected bandgap cell. *Proceedings. IEEE 38th Midwest Symposium on Circuits and Systems*, 1:pages 397–400. August 1995.

[22] Made Gunawan, Gerard C. M. Meijer, Jeroen Fonderie, and John H. Huijsing. A curvature-corrected low-voltage bandgap reference. *IEEE Journal of Solid-State Circuits*, SC-28(6):pages 667–670, June 1993.

[23] W. Timothy Holman. A new temperature compensation technique for bandgap voltage references. *IEEE International Symposium on Circuits and Systems*, 1:pages 385–388, May 1996.

[24] Gerard C. M. Meijer, Peter C. Schmale, and Klaas van Zalinge. A new curvature-corrected bandgap reference. *IEEE Journal of Solid-State Circuits*, SC-17(6):pages 1139–1143, December 1982.

[25] B. Gilbert. Translinear circuits: a proposed classification. *Electronics Letters*, 11(1):pages 14–16, 1975.

[26] B. Gilbert. Current-mode circuits from a translinear viewpoint : A tutorial. In C. Toumazou, F. J. Lidgey, and D. G. Haigh, editors, *Analog IC Design: The Current-Mode Approach*, chapter 2, pages 11–91. Peter Peregrinus for IEE, 1990.

[27] B. Gilbert. Translinear circuits: An historical overview. *Analog Integrated Circuits and Signal Processing*, 9(2):pages 95–118, 1996.

[28] E. Seevink. *Analysis and Synthesis of Translinear Integrated Circuits*. Elsevier, New York, 1988.

[29] Evert Seevinck and Remco J. Wiegerink. Generalized translinear circuit principle. *IEEE Journal of Solid-State Circuits*, SC-26(8):pages 1098–1102, August 1991.

[30] Remco J. Wiegerink. *Analysis and Synthesis of MOS Translinear Circuits*. Kluwer Academic Publishers, 1993.

[31] I. M. Filanovsky and Yiu Fai Chan. BiCMOS cascaded bandgap voltage reference. *Proceedings. IEEE 39th Midwest Symposium on Circuits and Systems*, 1:pages 943–946, August 1996.